



PHD

Exploring endosperm-led seed growth in *Arabidopsis thaliana*

Bouariky, Ahlam

Award date:
2010

Awarding institution:
University of Bath

[Link to publication](#)

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy

If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Exploring endosperm-led seed growth in ***Arabidopsis thaliana***

Copy 1 of 3

Ahlam Hamad Bouariky

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September, 2010

Attention is drawn to the fact that copyright of this thesis rests with its author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and they must not copy it or use material from it except as permitted by law or with the consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Signed:

Ahlam Bouariky

Dedicated to:

My father, husband and children

Acknowledgments

I am heartily thankful to my supervisor, Prof. Rod Scott, whose guidance, encouragement and support from the initial to the final level enabled me to develop an understanding of the subject. I am also very grateful to Dr. Sushma Tiwari, someone to whom I owe a real debt of gratitude. Your useful technical support, suggestions, encouragement and friendship are greatly appreciated.

Thank also go to Dr. Melissa Spielman for here technical support with microscopy and to all the past and present members of lab 3S 1.13 for their support and friendship: Rhiannon, Lucy, Valantyna, Bolbol, Alec and Holly. To my friends Dina and Amani thank you for your support and friendship and making my life easier abroad. I also thank Dr. Claudia Kohler (EZH) for providing *phe1.1* mutant seeds.

To my husband Meshal, and our children Fatema, Abdulla and Taiba, all I can say is it would take another thesis to express my deep love for you. Your patience, love and encouragement have support me, mostly in those many in which I spent more time in the lab or with my computer than with you.

I would like to thank the Department of Biology and Biochemistry at University of Bath for their support and providing a good training environment for students. Lastly, big thanks go to my sponsor Kuwait University for the financial support.

Abstract

The food supply to a growing world population is based on grain crops, which are harvested for their seeds. Therefore, understanding the process of seed development and its regulation has been important to increase production. This has become even more relevant and important as grain production reaches a plateau post green revolution and food security for an ever-increasing population becomes more pressing. This study focuses on altering seed size via manipulation of the endosperm, an important component of the seed that not only nourishes the developing embryo and provides the majority of our food. A classic study done by crossing different ploidies of *Arabidopsis thaliana* (*A. thaliana*) results in changes to endosperm-led growth as a consequence of altered parental genome ratios in the endosperm specifically. Increased paternal contribution results in an enlarged endosperm, and a heavier seed, whereas increased maternal contribution has the opposite effect.

Whole genome transcript profiling using microarray analysis of siliques generated by interploidy crosses identified *A. thaliana* genes namely: *PHE1*, *PHE2*, *AGL40*, *AGL62*, *AGL28*, *AGL45*, *CKX2*, *MAPK10*, *E2L2*, *GAI*, *CYCD4;1*, *CYCD4;2*, *GA20OX5*, *GA-regulated* and *AT5G46950*, that are positively associated with endosperm overgrowth. In order to verify the role of these genes in endosperm proliferation, knock in (KI) which causes gene over-expression and knock out (KO) which results in gene inactivation, strategies were used. Constitutive over-expression of *CKX2*, *MAPK10*, and *E2L2* showed abnormal vegetative and floral phenotypes, whereas targeted endosperm-specific over-expression of *PHE2* and *GAI* showed an increase in seed size and/or fertility. KI plants of *AGL40*, *AGL62*, *AGL45*, *CYCD4;1*, *CYCD4;2*, *GA20OX5*, *GA-regulated* and *AT5G46950* did not result in any obvious phenotypic effects under normal growth conditions. KO mutant plants namely: *phe1*, *phe2*, *agl40*, *agl62*, *agl28* and *agl45*, as single individual mutants were also indistinguishable to wild type plants in non-seed phenotypes. Embryo sac area, individual seed weight and total seed yield data obtained from a *phe1* line showed a smaller embryo sac area, lighter seed and reduced total seed yield. Loss of function of *agl62* showed precocious endosperm cellularisation and seed lethality. It is nearly impossible to generate *phe1/phe2* double mutants due to a very tight linkage between these genes (on the chromosome physically the two genes lie next to each other) and hence doubles, triples and quadruple mutants where appropriate were made with either *phe1* or *phe2* in

combination with other mutant lines of the MADS family of transcription factors. None of the double mutant combination tested had obvious developmental defects. However, the triple mutant *phe1/phe1::agl40/agl40 ::agl45/agl45* and the quadruple mutant *phe2/phe2::agl40/agl40::agl45/agl45 ::agl28/agl28* showed abnormal embryo and endosperm development which resulted to seed death. This indicates a functional redundancy among these MADS box genes. However, it is unclear how PHE1, PHE2, AGL40, AGL45 and AGL28 act to affect embryo and endosperm development. This result confirms that these genes are positively associated with endosperm over-proliferation as growth promoters and function in some cases as singly or as participants in a complex to control seed development. Thus manipulating expression of genes involved in endosperm development in *A. thaliana* and the potential use of this knowledge in crop plants provides us with a route to improvement of crop yields.

Abbreviations & Acronyms

AGL	AGAMOUS-LIKE
ANOVA	analysis of variance
Bp	base pair/s
BASTA	glufosinate ammonium
CaMV	cauliflower mosaic virus
Cds	coding sequence
CKX	cytokinin oxidase
CZE	chalazal endosperm
Col	Columbia
DAP	days after pollination
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
EDTA	ethylenediaminetetraacetic acid
E2L2	E2F-like repressor
F1	first filial generation
F2	second filial generation
FIE	FERTILIZATION INDEPENDENT ENDOSPERM
FIS1	FERTILIZATION INDEPENDENT SEED 1
Kb	kilo base pairs
GA	gibberellin
kD	kilo Dalton
KI	Knock in
KO	Knock out
<i>Ler</i>	Landsberg <i>erecta</i> (mutant)
Min	Minutes
MADS	MCM1-AGAMOUS-DEFICIENS-SRF
MEA	MEDEA
mRNA	messenger RNA
NASC	Nottingham Arabidopsis Stock Centre
OCS	Octopine synthase terminator
PCR	polymerase chain reaction
PcG	Polycomb-group proteins
PER	peripheral endosperm
PHE1	PHERES1
PHE2	PHERES2
<i>pPER</i>	At5g46950 endosperm-specific promoter
RNA	ribonucleic acid
RT-PCR	reverse transcriptase-PCR
Sec	seconds
S.E.M	standard error of the mean
TAE	tris-acetate-EDTA
Ws-4	Wassilewskija-4
WT	wild type

Contents

Acknowledgments	3
Abstract	4
Abbreviations & Acronyms	6
Chapter 1: General introduction	11
1.1 The importance of seed	11
1.2 The model plant <i>A. thaliana</i>	12
1.3 Seed development in <i>A. thaliana</i>	13
1.3.1 The life cycle of flowering plant	13
1.4 Endosperm	16
1.4.1 Origin and genomic composition of endosperm	16
1.4.2 Endosperm development	16
1.4.3 The role of endosperm in seed development	18
1.4.4 Molecular regulators of endosperm development	19
1.5 Parent-of-origin-effect on seed development	19
1.6 Microarray and transcript profiling	22
1.7 Project aim and objectives:	24
Chapter 2: Materials & Methods	26
2.1 Materials	26
2.1.1 Plant material	26
2.1.2 Primer design and synthesis	26
2.1.3 Sequencing services and data analyses	27
2.1.4 Statistical analyses	27
2.1.5 Image capture and processing	27
2.2 Methods	27
2.2.1 Plant growth conditions	27
2.2.2 Phenotypic characterization	28
2.2.3 Seed germination	28
2.2.4 Cross pollination	28
2.2.5 Restricted pollination	28
2.2.6 Mature seed collection, seed weight and total seed yield	28
2.2.7 Production of double, triple and quadruple T-DNA and transposon insertion lines	29
2.2.8 Selection of over-expressed homozygous single insertion lines	29
2.2.9 Seed clearing and differential contrast (DIC) microscopy	29
2.2.10 Feulgen staining and confocal microscopy	30
2.2.11 Plant DNA extraction	31
2.2.12 Plant RNA extraction	31
2.2.13 Amplification of DNA for genotyping and expression analyses	31
2.2.14 cDNA synthesis	32
2.2.15 Agarose gel electrophoresis	32
2.2.16 Genotyping T-DNA and transposon insertion lines	32
2.2.17 Gel purification of DNA fragments	33
Chapter 3: Verification of genes involved in endosperm growth and development using a knock-in (KI) strategy	36
3.1 Introduction	36
3.1.1 Genes associated with endosperm over-proliferation	37

3.2	Results	38
3.2.1	The selection of genes positively associated with endosperm proliferation	38
3.2.1.1	Transcription factor genes	39
3.2.1.2	Genes involved in cell cycle	40
3.2.1.3	Genes involved in cell signalling or hormonal pathways	42
3.2.1.4	Genes involved in other biological process	43
3.2.2	Phenotypic analysis of plants containing 35S promoter driven candidate genes	50
3.2.2.1	35S-KI plants of <i>E2L2</i> , <i>MAPK10</i> and <i>CKX2</i> resulted in vegetative abnormalities	50
3.2.3	<i>pPER</i> promoter	55
3.2.3.1	Analysis of <i>pPER::PHE2</i> plants	56
3.2.3.1.1	Seed size analysis	56
3.2.3.1.2	Expression of <i>PHE2</i> in <i>pPER::PHE2</i> plants	58
3.2.3.2	Analysis of <i>pPER::CYCD4;1</i> plants	60
3.2.3.2.1	Seed size analysis	60
3.2.3.2.2	Expression of <i>CYCD4;1</i> in <i>pPER::CYCD4;1</i> plants	62
3.2.3.3	Analysis of <i>pPER::GA1</i> plants	63
3.2.3.3.1	Seed size analysis	63
3.2.3.3.2	Expression of <i>GA1</i> in <i>pPER::GA1</i> plants	65
3.2.3.4	Seed size analysis of <i>pPER::GA20-Oxidase 5</i> plants	66
3.3	Discussion	68
3.3.1	Expression of candidate genes	68
3.3.2	Over-expression of the candidate genes from the 35S promoter	68
3.3.2.1	Over-expression of <i>E2L2</i> , <i>MAPK10</i> and <i>CKX2</i> under the 35S promoter results in vegetative abnormalities	69
3.3.3	Over-expression of the candidate genes by using <i>pPER</i> promoter	71
3.3.3.1	Over-expression of genes which positively associated with endosperm growth has the potential to increase seed size	71
3.3.4	Future work	73
Chapter 4:	Phenotypic characterization of <i>phe1</i>	74
4.1	Introduction	74
4.1.1	<i>PHE1</i> is a MADS box transcription factor	74
4.2	<i>PHE1</i> expression is associated with endosperm over-proliferation	74
4.2	Results	78
4.2.1	Verifying the KO status of the <i>phe1.1</i> mutant alleles	78
4.2.2	Seed size analyses of <i>phe1.1</i> plants	80
4.2.2.1	Embryo sac area measurement using pollination time and embryo stage	80
4.2.2.2	Individual seed weight and total yield analyses	83
4.2.2.2.1	Does the size difference between <i>phe1.1</i> and wild type seed enable visual identification?	84
4.2.2.3	Relationship between embryo sac area, seed mass and endosperm size in <i>phe1.1</i> and wild type	87
4.2.3	The role of <i>PHE1</i> in interploidy crosses	89
4.2.3.1	Seed weight and abortion frequency analyses of 4x <i>phe1.1</i> reciprocal interploidy crosses	89
4.2.3.2	Seed weight and abortion frequency analyses of a restricted (4x) <i>phe1.1</i> reciprocal interploidy crosses	92
4.3	Discussion	94
4.3.1.	<i>PHE1</i> loss-of-function overall vegetative phenotype appear like wild type	94

4.3.2	<i>PHE1</i> loss-of function seed phenotype	94
4.3.2.1	The <i>phe1.1</i> allele show a reduction in endosperm, individual seed size and total seed yield	94
4.3.2.1.1	The small seed size of <i>phe1.1</i> is apparent compared to wild type	95
4.3.2.2	The reduction in endosperm nuclei number is associated with the reduction in seed mass and endosperm size of <i>phe1.1</i>	95
4.3.3	Extra copy of paternal <i>PHE1</i> potentially contributes to seed death in interploidy crosses	96
4.3.4	Future work	98
Chapter 5: Identifying the role of <i>AGL62</i> in endosperm growth and development		99
5.1	Introduction	99
5.1.1	<i>AGL62</i> is a type 1 MADS box transcription factor associated with endosperm over-proliferation	99
5.2	Results	100
5.2.1	Analysing of <i>AGL62</i> loss-of-function phenotype	100
5.2.1.1	Genotyping <i>AGL62</i> T-DNA insertion line	100
5.2.1.2	Seed set of <i>AGL62/agl62</i> and wild type plants	102
5.2.1.3	<i>AGL62</i> segregation	103
5.2.1.4	Loss of <i>agl62</i> function results in early seed lethality	103
5.2.2	Analysis of <i>AGL62</i> over-expression lines	105
5.2.2.1	Screening and selection of <i>AGL62</i> single insertion lines	106
5.2.2.2	Seed size analysis of <i>pPER::AGL62</i> plants	106
5.3	Discussion	108
5.3.1	<i>AGL62</i> mutation affects seed development	108
5.3.2	<i>AGL62</i> controls early endosperm cellularisation and possibly will promote proliferation	108
5.3.3	Over-expression of <i>AGL62</i> in endosperm has no effect on final seed size	110
5.3.4	Future work	110
Chapter 6: Verification of genes involved in endosperm growth and development using a knock-Out (KO) strategy		112
6.1	Introduction	112
6.1.1	Genes associated with endosperm over-proliferation	112
6.2	Results	115
6.2.1	Genotyping T-DNA insertion lines	115
6.2.2	Sequencing T-DNA homozygous insertion lines	115
6.2.3	Embryo sac area measurements of the homozygous T-DNA insertion mutants	118
6.2.3.1	Seed size analyses of the single homozygous T-DNA insertion plants	118
6.2.3.1.1	Embryo sac area measurement	118
6.2.3.1.2	Individual seed weight analyses	124
6.2.4	Analysis of double mutants	125
6.2.4.1	Analyses of the <i>phe2/phe2::agl40/agl40</i> plants	126
6.2.4.2	Seed size analyses of <i>phe1/phe1::agl40/agl40</i> plants	128
6.2.5	Analysis of triple mutants	130
6.2.5.1	<i>phe2/phe2</i> triple homozygous complexes	130
6.2.5.1.1	Seed size analyses of <i>phe2/phe2::agl40/agl40::agl28/agl28</i> plants	131
6.2.5.1.2	Seed size analyses of <i>phe2/phe2::agl40/agl40::agl45/agl45</i> plants	132
6.2.5.2	<i>phe1/phe1</i> triple homozygous complexes	135

6.2.5.2.1	Seed size analyses of <i>phe1/phe1::agl40/agl40::agl28/agl28</i> plants	136
6.2.5.2.2	Seed phenotype analyses of <i>phe1/phe1::agl40/agl40::AGL45/agl45</i> plants	138
6.2.6	<i>phe2</i> -centered quadruple mutant	141
6.2.6.1	Seed phenotype analyses of <i>phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28</i> plants	142
6.3	Discussion	146
6.3.1	Homozygous T-DNA insertions did not affect the vegetative phase	146
6.3.2	Endosperm size and individual seed weight	146
6.3.2.1	The <i>phe2.1</i> , <i>agl40.1</i> , <i>agl45.1</i> and <i>agl28.1</i> mutants have no discernable effect on seed development	146
6.3.3	Double mutants do not impact on seed development	148
6.3.4	All triple mutants had no seed phenotype except the <i>phe1::agl40::agl45</i> complex	148
6.3.5	The <i>phe2</i> -centred quadruple mutant showed seed lethality	151
6.3.6	Future work	153
Chapter 7:	General discussion	155
7.1	The candidate genes are positively associated with endosperm proliferation	155
7.2	The potential of candidate genes to alter seed size	156
7.2.1	The outcomes of the KI strategy	156
7.2.2	The outcomes of the KO strategy	157
7.2.2.1	Loss of <i>PHE1</i> expression reduces endosperm growth	157
7.2.2.2	<i>AGL62</i> controls endosperm cellularisation	158
7.2.2.3	MADS box transcription factor genes showed functional redundancy in controlling endosperm development	159
References		162

Chapter 1: General introduction

1.1 The importance of seed

The worldwide population is estimated to reach 9.1 billion by 2050, 34% higher than today (FAO, 2009a). This means farmers will have to increase food production by 70% in order to feed this larger population. This will require, for example, increasing annual cereal production from the current annual production of 2.1 billion tonnes to about 3 billion tonnes per year (FAO, 2009a). Increased use of food crops for biofuel production could have serious implications for food security. Biofuel production, also based on agricultural products, increased more than threefold from 2000 to 2008 (FAO, 2009a). In 2007-08 about 10% of global production of coarse grains was used in ethanol production.

The *status quo*, without additional demands and competition, leaves 1.02 billion hungry and undernourished people in the world today; this represents the highest level of hungry people since 1970 (FAO, 2009b). That means that almost one in six people do not get enough food to lead a healthy and active life, making hunger and malnutrition a greater risk to health (WFP, 2009). These figures have prompted some analysts to suggest that a second Green Revolution is needed to once again increase agricultural productivity (Shah and Strong, 2000).

The Green Revolution refers to a series of research, development, and technology transfer schemes between 1943 and the late 1970s that dramatically increased agricultural output in many developing countries. In addition, the Green Revolution of the 1960s and 1970s, with its package of improved seeds, farm technology, better irrigation and chemical fertilizers, was highly successful at meeting its primary objective of increasing crop yields and boosting total food supplies (FAO, 2002; 2007). As an example, cereal production more than doubled in developing countries, such as India, Pakistan and Mexico, between the years 1960–2000. Yields of rice increased by 109%, maize by 157% and wheat by 208%. The production increases can be credited approximately equally to irrigation, fertilizer, and plant varieties (FAO, 2004).

The Green Revolution varieties of wheat and rice combined shortened stem length, which increased harvest index and allowed effective use of nitrogen fertilizer, with superior disease resistance. In the case of maize, improvements centred on tolerance to crowding which enabled higher planting densities and therefore greater yields/unit land area (Duvick,

1999, 2005). Increasing yield by these approaches is showing signs of reaching a plateau (Finger, 2010). Therefore new ways of increasing yield must be found to ensure that output meets increasing demand. The food supply for the growing world population is based on grain crops, which are harvested for their seeds. Understanding the process of seed development and its regulation is therefore potentially important in developing strategies to increase production. Since agriculture began, plants have been subjected to selection and breeding for size, and most food grains consumed today have seeds significantly larger than their wild relatives (Sundaresan, 2005). Seed contains the plant embryo and the endosperm, which sustains embryo growth. Endosperm represents 60% of world's food supply (Berger, 2003). Therefore, enhancing endosperm development is a potential way to increase total seed yield.

There is a trade-off between seed size and seed number due to the limited resources of the mother plant (Venable, 1992). Many examples of the negative correlation between seed size and number were reported in *Arabidopsis thaliana* (*A. thaliana*) (Jofuku *et al.*, 2005; Schruff *et al.*, 2006) and crop plants (Kiniry *et al.*, 1990). While seed number appears greatly influenced by environmental conditions and resource availability, seed size is primarily controlled by genetic factors and as a result is more stable (Sadras, 2007). A recent comparative study across crop species with respect to resource distribution and the trade-off between seed number and seed weight supported the trade-off theory (Gambin and Borrás, 2010). The author concluded that seed number and individual seed weight combinations across species were related and could be explained by considering resource availability when plants were able to adjust seed number to the growth. He added that the available resources around the seed set period are proportionally allocated to produce either many small seeds or few larger seeds depending on the particular species.

Identification of genes involved in endosperm development provides both improved understanding of the regulation of this crucial process and potential targets for manipulation of endosperm size. The project described in this thesis aimed to identify such genes in the model plant *A. thaliana* with the potential to engineer larger endosperms in crop plants, to therefore increase seed size and then total seed yield.

1.2 The model plant *A. thaliana*

A. thaliana is a small flowering plant that is widely used as a model organism in plant biology. *A. thaliana* is a member of the mustard (Brassicaceae) family, which includes

cultivated species such as oilseed rape, *Brassica napus*. Although of no direct economic importance *A. thaliana* has proved highly useful as a model system for studying numerous aspects of biology in flowering plants (Somerville and Koornneef, 2002). *A. thaliana* plants are small, with a rapid life cycle of about 6 weeks from germination to seed maturation and are therefore easily cultivated in restricted space. It is self-fertile with a prolific seed production and also the plant is easily cultivated in a restricted space. *A. thaliana* also offers several advantages for molecular analysis including its very small genome consisting of ~130 Mb, which contains ~ 25,000 genes. In contrast, crop species have large genomes such as *Brassica napus* (~1,200 Mb) and *Triticum aestivum* (~16,000 Mb), which makes molecular studies difficult (Arumuganathan and Earle, 1991). In addition, the *A. thaliana* has extensive genetic and physical maps of all five chromosomes which are available at the TAIR website (www.arabidopsis.org), which provides extensive information about genes, markers, polymorphisms, maps, clones, proteins and gene families. *A. thaliana* users also benefit from seed stock centres, such as the Nottingham Arabidopsis Stock Centre (NASC), that hold large collections of naturally occurring accessions (ecotypes) and mutant lines. A wide range collection of T-DNA insertion mutants is also available (Alonso *et al.*, 2003) allowing the role of many genes with unknown function to be identified. Other advantages of *A. thaliana* include the ease of transformation using *Agrobacterium tumefaciens* and efficient mutagenesis with the chemical mutagen (ethyl methanesulfonate (EMS)).

1.3 Seed development in *A. thaliana*

Seed development is essential to the reproductive success of flowering plants (Ungru *et al.*, 2008). Seeds are the dispersal units that allow plants to spread out and grow in new territories. They also help plants to survive under unfavourable conditions in a dormant stage. In addition, they have a nutritional value that supports the growth of the germinating seedlings. In the seed, the embryo, which represents the next plant generation, is enclosed by a nutritive tissue called endosperm; both the embryo and the endosperm are generated by the process of double fertilization that is unique to flowering plants.

1.3.1. The life cycle of flowering plant

In flowering plants, the diploid generation produces a special lineage leading to meiosis. The haploid meiotic products (spores) then develop as morphologically different life forms, the male and female gametophytes, which will differentiate a germ-line and produce

the male and female gametes (Figure 1.1) (Spielman *et al.*, 2001). In flowering plants, the pollen grain is the male gametophyte and the embryo sac is the female gametophyte. The female gametophyte is enclosed within an ovule situated in the ovary, which develops into the fruit after fertilization. Each embryo sac consists of a small number of different cell types; the egg and central (polar) cells are capable of fertilization, whereas the synergids and antipodals are sterile cells (Baroux *et al.*, 2002). Conversely, the male gametophyte ends its early development within the anther. Pollen mother cells undergo successive meiotic divisions to produce spores (pollen grains) consisting of three cells - the vegetative cell and the two sperm cells, which lie within the cytoplasm of the vegetative cell (Mascarenhas, 1989).

Double fertilization in flowering plants was independently discovered more than a century ago by Navashin (1898) in Russia and Guignard (1899) in France (Friedman, 1998). Double fertilization is a unique feature of flowering plants and consists of two important events that start the process of seed development (Figure 1.1). Once the pollen matures and germinates on a receptive stigma, the two sperm cells are delivered via the pollen tube into the embryo sac. The first event in double fertilisation occurs when one of the two sperm cells fuses with the egg cell, creating the diploid zygote ($2n$), while the second event is the fusion of the second sperm cell with the central cell giving rise to the usually triploid endosperm (Floyd and Friedman, 2000; Baroux *et al.*, 2002; Friedman and Williams, 2004; Berger *et al.*, 2008; Friedman *et al.*, 2008; Friedman and Ryerson, 2009).

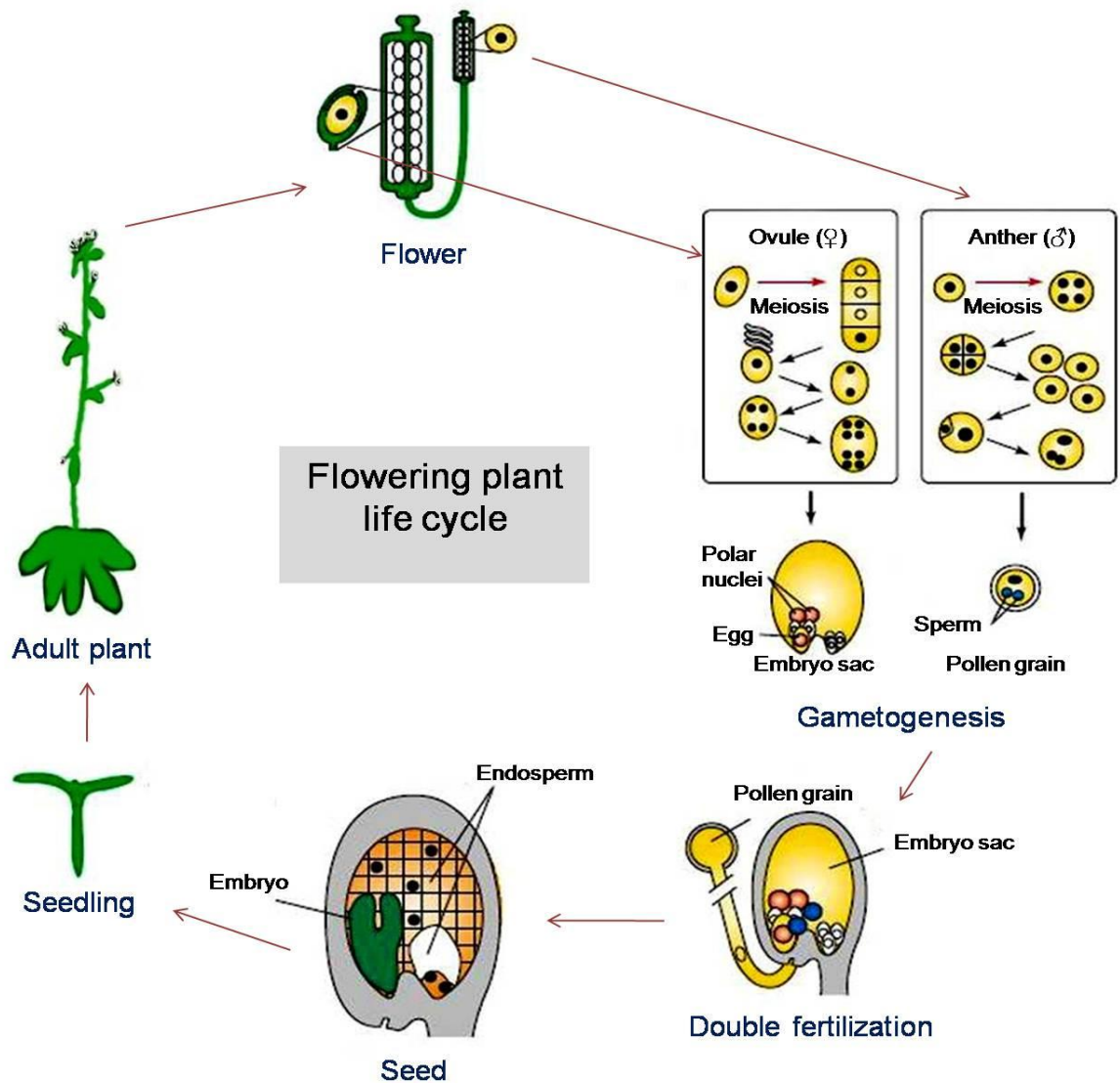


Figure 1.1 Life cycle of flowering plants (adapted from Spielman *et al.*, 2001). Gametogenesis and double fertilisation; the pollen grain transmits two sperm to the embryo sac, where one fertilizes the egg, giving rise to the embryo, and the other fertilizes the fused polar nuclei (central cell nucleus), producing the endosperm. Endosperm is a terminal tissue that acquires resources from the seed parent for use in embryo growth and/or germination. The germ-cell lineage does not differentiate until flower formation in the adult plant. Female meiosis occurs in ovules and male meiosis in anthers. In most species, there is only one surviving megaspore (black nucleus) from each female meiotic event. The megaspore divides mitotically to form the embryo sac, which contains two female gametes: the haploid egg, and the diploid central cell. Male meiosis produces microspores that give rise to pollen grains, each containing two sperm. Double fertilization consisted of major steps, (1) the attraction and arrest of the pollen tube, (2) the release of the two male gametes in the degenerated synergid, (3) the migration of male gametes to the two female gametes, (4) gamete recognition and fusion, (5) the fusion of the parental genetic material during karyogamy (fusion of pronuclei of two cells), (6) re-initiation of the cell cycle, transcription and translation, leading to a new start of the zygotic life (reviewed by Berger *et al.*, 2008).

1.4 Endosperm

The endosperm is a nutrient sink that is supplied by the seed parent, stores resources, and nourishes the embryo during embryogenesis or germination. Successful embryogenesis therefore requires the development of the second fertilization product, the triploid endosperm (Lopes and Larkins, 1993).

1.4.1 Origin and genomic composition of endosperm

In the seed, the embryo is associated with endosperm that stores nutrients such as carbohydrates, oil and protein, that are essential to the survival of the next plant generation and an important source of nutrition in human and animal diets (Focks and Benning, 1998).

As detailed in (Section 1.3.1) the endosperm is the product of double fertilization, a unique biological process of flowering plants (Lopes and Larkins, 1993). One sperm cell (1 paternal; 1p) fuses with the egg cell (1 maternal; 1m), creating the diploid zygote (2n; 1m:1p genome ratio), while the other sperm cell fuses with the central cell giving rise to endosperm that is usually triploid (3n; 2m:1p genome ratio).

1.4.2 Endosperm development

A. thaliana has the nuclear type endosperm (Brown *et al.*, 1999; Brown *et al.*, 2003; Olsen, 2004) in which the endosperm develops through two stages, first syncytial and then cellular. Successive divisions of the triploid nucleus without cytokinesis (no cell plate formation) initiate the syncytial stage. This developmental program is proposed to allow the endosperm to grow more rapidly than the embryo, which it will sustain (Lopes and Larkins, 1993). The cytoplasm is then divided by cytokinesis in a process called cellularisation. Cellularisation of the endosperm is initiated in the area surrounding the embryo and proceeds through the central chamber to the chalazal pole, which remains syncytial until well after the rest of the endosperm has cellularised (Brown *et al.*, 1999). The different ‘bipolar’ developmental origins of chalazal and micropylar domains are a common pattern among the endosperms of all basal flowering plants (Floyd and Friedman, 2000). Since the bipolar endosperm development pattern of most flowering plants is shared with the bipolar pattern of embryo development, interactions between endosperm–embryo signalling is been suggested. *A. thaliana* is typical in having bipolar endosperm development in which a multinucleate chalazal region develops and remains uncellularised whilst simultaneously, the remnants of the endosperm, including the micropylar domain,

undergo cellularisation (Figure 1.2) (Kranz and Kumlehn, 1999; Nguyen *et al.*, 2000; Berger, 2003; Olsen, 2004; Berger *et al.*, 2006).

The endosperm of *A. thaliana* has three regions that become distinct as the seed matures and these regions are the embryo surrounding region or micropylar endosperm (MCE), peripheral endosperm (PEN) which lines the wall of the developing embryo sac, and chalazal endosperm (CZE), which develops adjacent to the vascular connection with maternal tissue (Olsen, 2004). Brown and co-workers 1999, studied endosperm cellularisation and reported that when the embryo attains the heart stage of development, cellularisation of endosperm initiates at the micropylar region and moves toward the central chamber of the embryo sac. In contrast, the chalazal region remains syncytial and forms nodules until the late stages of seed maturity.

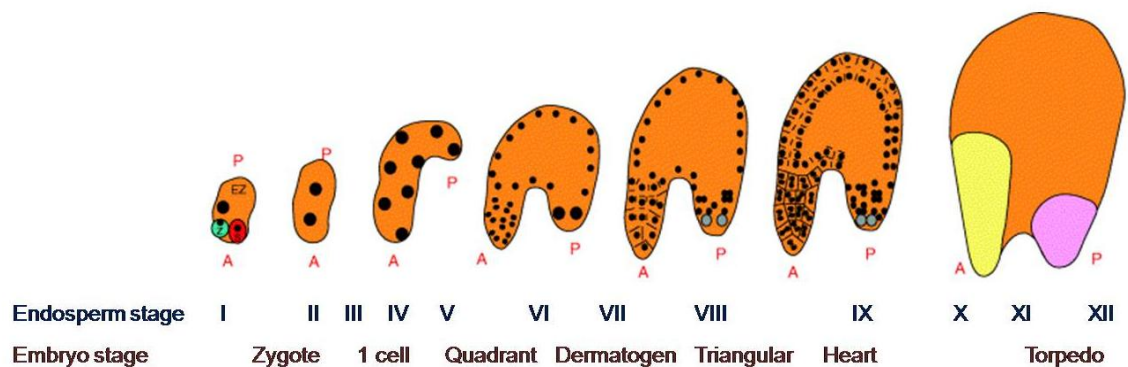


Figure 1.2 Major steps of endosperm development in *A. thaliana* (adapted from Berger, 2003). Double fertilization of the embryo sac by male gametes derived to the synergids produces two zygotic products: a true zygote from the fertilized egg cell (Z) and the endosperm zygote from the fertilized diploid central cell (EZ). Endosperm development is divided in two major phases, first the syncytial and then the cellular phase. Stages of development are defined by successions of pseudo synchronous mitosis. Three mitotic domains are defined from the anterior pole (A) to the posterior pole (P). The domains are micropylar endosperm (yellow), peripheral endosperm (orange) and chalazal endosperm (pink). The eighth mitotic cycle is followed by cellularisation of the syncytial peripheral endosperm.

1.4.3 The role of endosperm in seed development

Endosperm development is a critical component of seed development. A genetic and physiological balance between the endosperm, the embryo, and the maternal tissues has long been known to be important for successful seed development. In *A. thaliana* and most other flowering plants, the endosperm stores nutrients for consumption by the developing embryo during embryogenesis. However, other suggested roles of endosperm include regulation of seed size and fruit development (Grini *et al.*, 2002).

In cereals, endosperm plays a crucial role in maintaining the high osmotic potential around the embryo, provides a mechanical support during early embryo growth, and stores reserves, nutrients, and hormones to facilitate and support seed germination (Lopes and Larkins, 1993). In contrast, the endosperm of many dicotyledonous plants, including *A. thaliana*, is ephemeral being consumed almost entirely by the completion of seed development. In such species the cotyledons form the primary storage organ, but the endosperm has a role in supporting the early growth of the embryo until sufficient reserves are stored in the cotyledons (Marinos, 1970; Raghavan, 1986). Berger *et al.* (2006) reviewed the importance of endosperm as a source of nutrients for the growing embryo and as an integrator of seed growth and development by sharing signalling routes with the embryo and seed integuments. Analysis of an *A. thaliana* mutant of the cyclin dependent kinase *cdc2a*; *CDKA;1* provides a good example of the communication between the embryo and the endosperm during seed development (Nowack *et al.*, 2006). *cdc2a* (or *cdka;1*) mutant pollen has only one sperm cell (Iwakawa *et al.*, 2006; Nowack *et al.*, 2006) that exclusively fertilizes the egg cell to trigger embryo development up to the globular stage. Interestingly, the unfertilized endosperm nevertheless completed several rounds of the cell cycle leading to the suggestion that fertilization of the egg cell produces a positive signal that initiates proliferation of the central cell.

A further role proposed for the endosperm in seed development is to participate in communication between the endosperm tissue and maternal tissue to regulate the development of each other (Lopes and Larkins, 1993). For example, the maize *miniature-1*(*mn1*) mutant illustrated the role of the basal endosperm cells and their interaction with the adjacent maternal tissue. The mutant has a dramatically reduced in endosperm growth which was the cause of arresting embryo growth and production of small seeds (Miller and Chourey, 1992; Cheng *et al.*, 1996).

1.4.4 Molecular regulators of endosperm development

As with most developmental processes, the action of molecular regulators in endosperm development has been shown to play a key role in determining seed size (Sabelli and Larkins, 2009; Linkies *et al.*, 2010). A number of genes control endosperm development and function. In *A. thaliana*, a group of female-gametophyte expressed Polycomb group (PcG) 'FIS' class genes (*FERTILIZATION INDEPENDENT SEED 1 (FIS1)/MEDEA (MEA)*, *FIS2*, *FIS3/FERTILIZATION INDEPENDENT ENDOSPERM (FIE)*, and *MULTICOPY SUPPRESSOR OF IRA 1 (MSI1)*), strongly influence embryo and endosperm development, by for example, suppressing endosperm development in the absence of fertilization (Chaudhury *et al.*, 1997; Grossniklaus *et al.*, 1998; Kinoshita *et al.*, 1999; Kiyosue *et al.*, 1999; Ohad *et al.*, 1999; Kohler *et al.*, 2003a; Guitton *et al.*, 2004). The products of these genes are similar to the PcG proteins that regulate expression of genes through epigenetic silencing in *Drosophila* (Ohad *et al.*, 1999; Sorensen *et al.*, 2001). *MEDEA* was shown to regulate seed development by controlling the expression of the MADS box gene *PHERES1 (PHE1)* (Kohler *et al.*, 2003b).

Studies in *A. thaliana* have also identified sporophytic mutant alleles of three genes, *HAIKUI (IKU1)*, *IKU2* and *MINISEED3 (MINI3)* that produce a small seed associated with reduced endosperm growth and precocious endosperm cellularisation (Garcia *et al.*, 2005; Luo *et al.*, 2005). The *APETALA2 (AP2)* transcription factor can also alter seed size maternally, with *ap2* mutants producing larger seeds (Jofuku *et al.*, 2005; Ohto *et al.*, 2005). Recently, *AP2* was reported to have a significant effect on endosperm development: *ap2* mutant seeds undergo an extended period of rapid endosperm growth early in development, which was linked delayed endosperm cellularisation and overgrowth of endosperm central vacuole (Ohto *et al.*, 2009).

1.5 Parent-of-origin-effect on seed development

Seed development in many flowering plants depends on a correct ratio of (2m:1p) genomes in the endosperm and disturbance of this ratio leads to abnormal seed development (Lin, 1984; Scott *et al.*, 1998). Scott *et al.* (1998) carried out the first developmental study of seeds from interploidy crosses in *A. thaliana*. Seeds from ♀2x X ♂4x and ♀4x X ♂2x crosses are viable but differ in size according to their endosperm ratio 2maternal :2paternal (2m:2p) and 4m:1p respectively. However, an extra maternal genome contribution in seeds from the ♀4x X ♂2x cross resulted in hypo-proliferated endosperm and a small seed. On the other hand, a ♀2x X ♂4x cross that affords an extra paternal genome in the

endosperm (5x) produced an over-proliferated endosperm and a large seed. Seeds from the ♀2x X ♂6x and ♀6x X ♂2x crosses were aborted seeds. This result suggested that altering the maternal and paternal genomes contribution in *A. thaliana* seeds leads to dramatic effects on seed size and viability.

The visible differences following interploidy crosses are often much more dramatic in the endosperm than in embryo itself (Scott *et al.*, 1998). The events within the endosperm resulting from parental genome imbalance were that an extra paternal genome (growth promoters) was associated with accelerated mitosis and delayed cellularisation, while seeds with an extra maternal genome (growth suppressors) show reduction in endosperm mitosis and precocious cellularisation that produce small plump seeds. This implies that growth-promoting genes are mainly expressed from the paternally inherited genome and are silent in the maternally inherited genome. Conversely, growth-suppressing genes are expected to be primarily maternal expressed and paternal silent (Kohler *et al.*, 2005). This data is consistent with the parental conflict theory that the mother distributes resources to her offspring from different fathers equally, whereas the father desires his progeny to grow faster and obtain more maternal resources than the progeny of other fathers. In flowering plants, the conflict is suggested to act on endosperm growth and development and only indirectly affect the embryo (Haig and Westoby, 1991).

A molecular explanation of the parent of origin effect is provided by genomic imprinting which is a type of epigenetic gene regulation where the expression of a gene depends on the sex of the parent from which it is originated (Haig and Westoby, 1991; Grossniklaus *et al.*, 1998; Spielman *et al.*, 2001). Lin (1984) explained the 2m:1p requirement as an indication that parentally imprinted genes are involved in development of the endosperm. Subsequent genetic and molecular evidence indicated that in flowering plants imprinted genes function primarily in the endosperm because the embryo could complete the development with a 1m:0p or 2m:0p genomes constitution, whereas maternal and paternal genomes contributions were needed in the endosperm of all sexually producing flowering plants (Kermicle and Alleman, 1990).

Imprinting in flowering plants had been hypothesized to require DNA methylation (Adams *et al.*, 2000). Therefore, it was thought possible to change the parent-of-origin effect in endosperm growth and development by altering the DNA methylation levels in the loci of the imprinted gene (Adams *et al.*, 2000). Studies were reported on crosses within *A. thaliana* using an antisense gene (*MET1a/s*) targeting the major cytosine DNA

methyltransferase gene *DNA METHYLTRANSFERASE 1 (MET1)* which reduces cytosine methylation by as much as 85% (Ronemus *et al.*, 1996). Crosses between one parent that had a reduced level of DNA methylation due to the *MET1a/s* gene and a wild type parent of the same ploidy produced a seed phenotype very similar to interploidy crosses with respect to characteristics diagnostic for parental genomic imbalance; if the hypomethylated plant was the seed parent the endosperm was paternalized and seeds were large and heavy, while if the pollen parent was hypomethylated the endosperm were maternalized and seeds were small and light (Adams *et al.*, 2000). These phenotypes indicated that DNA hypomethylation has a profound effect on endosperm development depending on the parental genome that was hypomethylated (Adams *et al.*, 2000; Vinkenoog *et al.*, 2000; Spielman *et al.*, 2001; Vinkenoog and Scott, 2001). It is concluded that DNA methylation acts to silence maternally-inherited endosperm-promoting genes and paternally-inherited endosperm-inhibiting genes.

As reviewed in Spielman *et al.* (2001), experiments that combined hypomethylation and *fis*-class mutants provide further evidence that methylation of the paternally-inherited genome silences endosperm inhibiting genes. As mentioned in Section 1.4.4, the maternal inheritance of *fis*-class mutant alleles resulted in endosperm over-proliferation and seed abortion. For example a cross between *mea* mutant as a seed parent and a hypomethylated *ddm1* (*decrease in dna methylation 1*) mutant as a pollen parent will rescue seed abortion (Vielle-Calzada *et al.*, 1999). Also, crosses between *fie* mutants (seed parent) with *MET1a/s* (pollen parent) showed similar phenotype that rescued seed abortion (Vinkenoog *et al.*, 2000). All *fis*-class mutants resulted an endosperm phenotype very similar to that generated by interploidy cross with a paternal excess genome (e.g. 2x X 6x) suggesting that seed abortion of these mutants is caused by the released silencing of maternally-inherited endosperm-promoting genes (Spielman *et al.*, 2001).

The parental conflict hypothesis also suggests that genomic imprinting developed as a consequence of conflict between maternally- and paternally-inherited genomes over resource allocation from mother to offspring as mentioned above. In addition, the opposite seed phenotypes produced from interploidy crosses (Scott *et al.*, 1998) provided strong evidence that the maternally- and paternally-inherited genomes are not alike and have aggressive effects on endosperm development. In support of the parental conflict theory these results indicate that paternally-expressed imprinting genes promote endosperm proliferation, whereas maternally-inherited imprinted genes restrict this growth.

1.6 Microarray and transcript profiling

Interploidy crosses cause large changes in endosperm development due to the action of imprinted genes. Whilst there is a fairly large body of knowledge about the regulation of imprinting (Section 1.5), much less is known about the genes that are targeted by imprinting, either directly or indirectly, to bring about the observed changes in endosperm development. At the time of starting the work reported in thesis there were only four imprinted genes described in *A. thaliana*. The Scott laboratory therefore set out to identify further imprinted genes using a microarray approach.

The availability of microarrays representing almost all the genes present within the *A. thaliana* genome allow visualisation of the entire transcriptome of specific tissues and organs. Conducting transcript profiling on seeds that had been manipulated in interploidy crosses to produce larger or smaller endosperm (Figure 1.3) provided an insight into the basis of the interploidy cross effect on seed and a means to identify the genes that are potentially useful to alter the endosperm size and then increase seed size and yield in crop plants.

The Scott laboratory had previously used a microarray approach to analyse the expression of seed genes in different interploidy crosses relative to the balanced 2x X 2x cross (Tiwari *et al.*, 2010). They had been investigated gene expression underlying paternal and maternal excess phenotypes using an Affymetrix microarray platform. The RNA samples were extracted from 5 DAP siliques obtained from interploidy crosses; 6x X 2x, 4x X 2x (maternal excess) and 2x X 4x, 2x X 6x (paternal excess) and balanced cross (2x X 2x) then were hybridized to Affymetrix ATH1 full-genome chips. The microarray analysis of interploidy crosses yielded a list of 20,442 genes and, after subtracting all genes either below a threshold of absolute expression or which showed unreliable expression from the data set, the final list included 14,944 genes that were present at least in one sample (Tiwari *et al.*, 2010). The data set identifies genes whose transcript abundance either increase, decrease, or remain unchanged relative to normal cross (2x X 2x) seeds of the same age. For example, in the ($\text{♀}2\text{x} \times \text{♂}4\text{x}$) cross, where the endosperm over-proliferates, we expected to find an increase in the expression level of growth promoters, and in ($\text{♀}4\text{x} \times \text{♂}2\text{x}$) we expected a high expression of growth inhibitor genes. Comparing the behaviour of genes across multiple data sets give a shortlist of genes either positively or negatively associated with endosperm-led seed growth. Tiwari *et al.* (2010) reported 114 genes in total that were associated with endosperm over-proliferation: their expression was up-

regulated in paternal excess crosses (2x X 4x and 2x X 6x) and not up in maternal excess crosses (4x X 2x and 6x X 2x). In addition the authors reported 119 genes associated with endosperm under-proliferation: their expression was up-regulated in 4x X 2x and 6x X 2x crosses (maternal excess) and not in paternal excess crosses. From the genes that were associated with endosperm over-proliferation, a shortlist was created that provided the candidate genes for study reported this thesis. These are described in more detail in Chapter 3 (3.2.1).

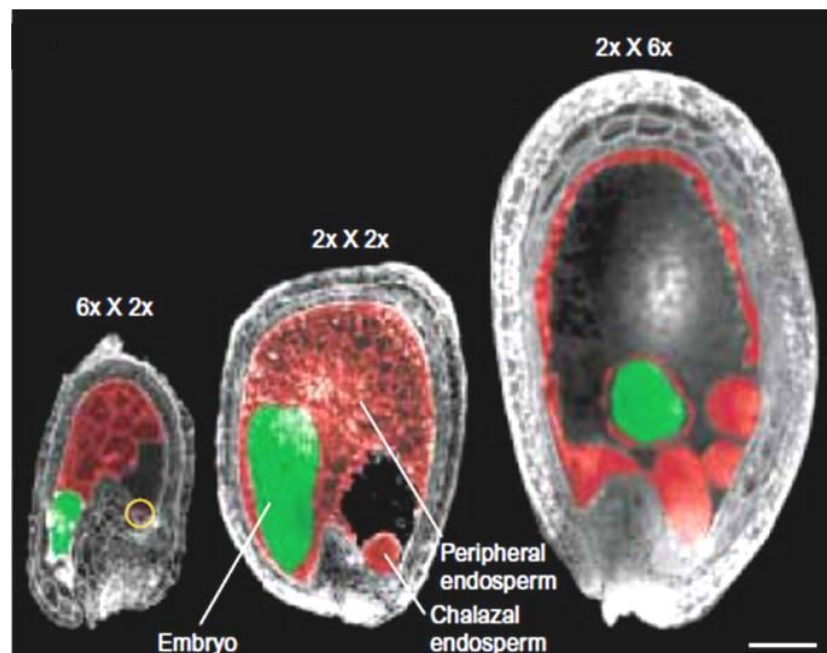


Figure 1.3 Seed phenotype of wild type and seeds produced from interploidy crosses (adapted from Spielman *et al.*, 2001). Confocal micrographs showing seeds produced by wild type *A. thaliana* plants of different ploidies. Embryos and endosperms are artificially coloured green and red, respectively. Balance of maternal and paternal genomes is altered by interploidy crosses. A normal seed from a 2x X 2x cross (middle) is compared with a maternalized seed (6x X 2x, left) and a paternalized seed (2x X 6x, right), all at 8 DAP. The maternalized seed is small with a reduced peripheral and chalazal endosperm (yellow circle). The paternalized seed is large with overgrown peripheral and chalazal endosperm, and endosperm does not cellularise. Scale bar is 100 μ m.

In addition to the Scott Laboratory microarray analysis of seed manipulated by interploidy crosses, other laboratories have reported microarray analyses on the seed of *A. thaliana*. A microarray analysis performed jointly in Goldberg Laboratory (UCLA, USA) and the

Harada Laboratory (UC Davis, USA), used *A. thaliana* Affymetrix Gene Chips, Laser Capture Microdissection (LCM), and 454 high-throughput sequencing technologies to profile the mRNA sets present in different seed regions and compartments throughout seed development (<http://seedgenenetwork.net/>). Tiwari *et al.* (2010) reported that there was 75% overlap between their dataset and Goldberg and Harada dataset of the genes that were up-regulated in peripheral endosperm.

Day *et al.* (2008) also generated microarray data set for developing *A. thaliana* seed by hybridizing LCM endosperm-derived target alongside target from similarly treated silique tissues using a two-colour microarray approach on 4 DAP silique. The authors identified 2,568 individual loci as being preferentially expressed in the endosperm, and a total of 793 genes of these genes exhibited early seed-specific and endosperm-preferred expression. Again there was considerable overlap (74%) between this data set and that of Tiwari *et al.* (2010).

In addition to the microarray analyses done specifically on seeds, other array data sets have been compiled to enable the visualisation of numerous cell, tissue and organ transcriptomes using such tools as GENEVESTIGATOR (www.genevestigator.com).

1.7 Project aim and objectives

The aim of this study was to identify genes with the potential to engineer endosperm growth and development and therefore to increase seed size and yield in crop plants. The objectives used to achieve this aim are summarised below.

1. Generate a shortlist of genes associated with endosperm over-proliferation from existing microarray data produced in the Scott laboratory (Tiwari *et al.*, 2010).

From the provided list of genes that were associated with endosperm over-proliferation the selection was made based on:

- Genes only up-regulated in paternal excess crosses and not maternal excess crosses.
- The expression profile within endosperm tissue by *in silico* analysis
- Involvement of the genes in an important biological function such as their participation in transcriptional control, cell cycle, regulating hormonal pathways, and others.

2. Determine using molecular genetics approach such as knock-In (KI) strategy whether the selected genes are capable of promoting growth and development

- In order to understand the role of the genes in endosperm development one of the complimentary experimental strategies proposed was KI: this provides gene over-expression by introducing extra copies of target gene into the plant genome. KI lines have an extra copy of the target gene expressed either from the 35S promoter, which results in vegetative over-expression, or an endosperm specific promoter (*pPER*), which results in endosperm over-expression. Since KI using the 35S promoter is likely to result in vegetative phenotypes (Sunilkumar *et al.*, 2002), over-expression of growth promoters should give large vegetative and reproductive organs. In contrast, KI using endosperm specific promoter is likely to result in seed phenotypes because the *pPER* is expressed in endosperm (Tiwari *et al.*, 2006). Therefore, over-expression of growth promoters should give large seed.

3. Determine whether the selected genes are capable of promoting growth and development by using a complimentary molecular genetics approach: Knock-Out (KO) strategy

- In order to understand the role of the genes in endosperm development, another complimentary experimental strategy was proposed. KO provides for gene inactivation by inserting a foreign DNA fragment within the target gene. Commonly used KO lines have T-DNA or transposon insertions, which disrupt the DNA coding sequence of the target gene, causing gene inactivation. KO should give much more specific effects, since the gene will be inactivated only where normally expressed. Therefore, KO of growth promoters should result in a smaller endosperm and seed size, assuming our target genes are mainly expressed in endosperm.

Chapter 2: Materials & Methods

2.1 Materials

2.1.1 Plant material

Seed stocks of Col-0, *Ler* and Ws-4 ecotypes were obtained from the Nottingham Arabidopsis Stock Centre (NASC), UK. *Arabidopsis thaliana* (*A. thaliana*) T-DNA insertion lines are as follow: Salk_105945, SALK_116450, SALK_022148, SALK_107011 (Col background) and FLAG_386G01 (Ws-4 background) T-DNA insertion lines were sourced from the Salk Institute Genomic Analysis Laboratory (SIGnAL) and obtained from NASC. The *phe1.1* (*Ler* background), which is a Ds transposon mutant was kindly donated by Dr. Claudia Kohler (ETH Zurich). Seeds of tetraploid (4x) wild type *Ler* and *phe1.1* were kindly donated by Prof. Rod Scott (University of Bath). The *A. thaliana* mutants that were used as an over-expressed insertion lines (*PHE2*, *AGL40*, *GA2OX2*, *AT5G46950*, *CKX2*, *MAPK10* and *E2L2*) were derived from Col-0 ecotype and obtained from (Ceres Inc, California, USA) as T2 seeds which were generated by constitutively expressing the candidate genes under a 35S promoter and have the *bar* gene conferring resistance to the herbicide BASTA as a selection marker. Another *A. thaliana* over-expressed lines (*PHE2*, *AGL40*, *AGL45*, *AGL62*, *GAI*, *CYCD4;1*, *CYCD4;2*, *GA20OX5*, *GA-regulated*, *AT5G46950*) were derived from Col-0 ecotype and provided by Dr. Sushma Tiwari (University of Bath) as T0 and T1 seeds which were generated by expressing the candidate genes under *pPER* promoter (Tiwari *et al*; 2006) and have the *bar* gene conferring resistance to the herbicide BASTA as a selection marker.

2.1.2 Primer design and synthesis

Primers were designed and checked by using NetPrimer (Premier Biosoft International), SIGnAL (www.signal.salk.edu/cgi-bin/tdnaexpress), TAIR (www.arabidopsis.org), and GeneDoc version 2.6.02 programs. After primer design, a megaBLAST (NCBI) search for highly related nucleotide sequences across the entire *Arabidopsis* genome was carried out to confirm target specificity. Primers were ordered from Invitrogen (Invitrogen, UK) and the information is shown in Table 2.1.

2.1.3 Sequencing services and data analyses

Sequencing of the KO mutants was carried out by cogenics (www.cogenics.com) or Geneservice (www.geneservice.co.uk). The sequence was viewed and analysed by Sequencher 4.9 Demo (www.genecodes.com).

2.1.4 Statistical analyses

Statistical tests were performed using Minitab 15 software (Minitab Inc, USA). Comparisons for normally distributed data with equal variances at 95% confidence level were carried out using the Student's *t*-test or one-way analysis of variance (ANOVA) followed by Tukey's multiple comparisons (family error rate, 5). Data which were not normally distributed was analyzed using the Mann-Whitney *U*-test.

2.1.5 Image capture and processing

Photographs of the whole plants and inflorescences were obtained using a Coolpix 4500 digital camera (Nikon). Dry seeds were captured under a SMZ1500 dissecting microscope (Nikon) using a Digital Sight DS-U1 colour camera (Nikon). Images were processed using Adobe Photoshop Elements.

2.2 Methods

2.2.1 Plant growth conditions

Seeds were stratified in 0.15% agarose electrophoresis grade (Invitrogen, UK), incubated at 4°C for 3–5 days. Stratified seeds were sown in Levingtons F2 compost treated with 0.2 g/l insecticide solution (Intercept 70 WG – Scotts, UK). Plants were grown in a Sanyo controlled environment rooms with 16/8 hours day/ night cycle, 70% relative humidity, 22°C (day) and 18°C (night), or in a glasshouse at 24°C (day) and 17.5°C (night).

For BASTA selection, seedlings were sprayed with a 427.5 µl/l of BASTA solution (150 g/l glufosinate-ammonium, AgrEvo) until completely covered on two separate occasions with a one week interval. The seedlings containing the transgene were BASTA resistant and therefore remained green. These selected green seedlings were grown for further phenotype analysis. For total seed yield analysis, plants were grown in 3 inch pots and fitted with an Aracon (Lehle Seeds) when primary inflorescence was 10 cm high.

2.2.2 Phenotypic characterization

KI and KO plants were examined for any differences in phenotypic characters such as leaf, stem, flower, siliques and seed and compared with the wild type phenotypic characters.

2.2.3 Seed germination

Seeds (30-40) were sown in a plastic Petri dish (90 mm) on sterile water saturated glass microfiber filter paper (70 mm, Whatman) in duplicates, wrapped with parafilm to reduce water evaporation. To test seed germination, plates were incubated at 22°C with a day length of 16 hrs; 18°C night and 70% humidity in a growth room for 15 days.

2.2.4 Cross pollination

Anthers of the seed parent were dissected from floral buds one day prior to anthesis. The flowers and siliques around the emasculated floral bud were removed to avoid any confusion with the self fertilized siliques. Two to three days after emasculation, mature pollen from the pollen parent was gently dabbed onto the mature stigma. Both emasculations and pollinations carried out under dissecting microscope (Leica MZ6).

2.2.5 Restricted pollination

Three floral buds on the primary inflorescence of six plants (seed parent) were emasculated, and all other buds and the inflorescence meristem were removed. The other branches of the plant were allowed to develop normally.

2.2.6 Mature seed collection, seed weight and total seed yield

Mature seeds were collected when pods were desiccated, and were stored in 1.5 ml Eppendorf tubes. The lid pierced to allow escape of any residual moisture. Tubes containing seeds were stored in a container with silica gel (Sigma) for further drying. Mean seed weight was measured using an UMT2 microbalance (Mettler-Toledo, Switzerland). To measure total yield, plants were grown using the Arasystem (Lehle Seeds, Tuscan, AZ), a plastic tube and cylinder support system that facilitates seed harvesting, ensuring collection of all seeds produced by an individual plant. Mature seed were separated from dry aerial organs using a sieve and stored in 1.5 ml tubes with a pierced lid and stored in a box containing silica gel to dry. The total seed yield was measured using a Sartorius A200S electronic balance (Sartorius Analytic).

2.2.7 Production of double, triple and quadruple T-DNA and transposon insertion lines

PHE2 and *PHE1* are both located on chromosome 1 (at 24255 kb and 24267, respectively). The production of *phe2/phe1* double mutant relied on a crossing over and recombination events between these loci, for that reason it was not impossible to generate *phe2/phe1* mutant but unrealistic due to the large number of F2 that would have to be screened because of the close linkage between the two loci. Therefore, the double, triple and quadruple were separated into two complexes. For example, *phe2/agl40* homozygous double mutants were produced by crossing *phe2* and *agl40* homozygotes to produce an F1 generation heterozygous for both T-DNA insertions. F2 plants were genotyped using PCR and the double homozygous mutants were selected for further study. The same procedure was carried out to produce *phe1/agl40* homozygous double mutants. The same procedure was used to generate triple and quadruple complexes centred around *phe1* and *phe2*.

2.2.8 Selection of over-expressed homozygous single insertion lines

Seeds from the T2 generation from each independent positive transformation line were screened for BASTA resistance. After spraying with BASTA, lines that showed 75% resistant seedling were believed to have a single copy of the transgene (single insertion line). Several resistant plants from this screen were allowed to set seeds. The T3 seed generation was then treated with BASTA and lines showing 100% resistance were considered to be a homozygous for the transgene. T3 single insertion homozygote plants were used for all further investigations.

2.2.9 Seed clearing and differential contrast (DIC) microscopy

In order to measure the embryo sac area of the ovules, siliques were first opened under a dissecting microscope using a sharp needle. The exposed ovules were picked up and mounted into 2-3 drops of the clearing solution (Chloral hydrate: H₂O: Glycerol; 8: 3: 1) on a new slide, overlaid with a cover slip which was then sealed by nail polish. To avoid squashing older and larger ovules (3 DAP and above), a bridge consisting of two cover slips was employed to support the sample cover slip. Samples were then kept at room temperature for different times depending on the developmental stages of the ovules, younger stages cleared faster. Cleared ovules were imaged using DIC field by 90i ECLIPSE microscope (Nikon) and photographed by either a Digital Sight DS-U1 colour camera (Nikon) or a Digital Sight DS-1QM/H black and white camera (Nikon).

2.2.10 Feulgen staining and confocal microscopy

Specimens were prepared using a protocol adapted from (Braselton *et al.*, 1996). All steps were carried out in 1.5 ml Eppendorf tubes.

Fixation: *Arabidopsis* siliques (2-8 DAP) were cut at both ends using razor blade, placed in a freshly prepared fixative (3:1, ethanol: acetic acid) and stored overnight at 4°C. Siliques were transferred to 70% ethanol and stored at 4°C, where they could be kept for several weeks until use or stored at -20°C for longer periods.

Staining: fixed siliques were washed three times in MilliQ H₂O for 15 min at room temperature, then hydrolyzed in 5 N HCL for 1 hour. The siliques were then washed three times for 10 min in MilliQ H₂O. Following the final wash, H₂O was replaced Schiff's reagent (Sigma) and siliques were kept for 2-3 hours in the dark for staining. The siliques were washed in 3x10 min changes of cold tap water.

Embedding: Following the staining step, siliques were washed once in 70% ethanol, once in 95% ethanol and twice in 100% ethanol, each for 10 min then stored at 4°C overnight. The 100% ethanol was changed every 1 hour until it stopped turning pink. At this point, materials could be stored in ethanol at -20°C for several weeks or processed as follows. Siliques were transferred to 1:1, ethanol:LR WhiteTM Resin (Agar Scientific) for 1 hour with regular mixing. The siliques were transferred to a pure resin for 1 hour, and then fresh pure resin was added before storing overnight at 4°C. A small drop of resin containing LR WhiteTM accelerator (Agar Scientific) was placed onto the middle of slide and two cover slips were pushed partway into it from either side. Under a dissecting microscope (Leica MZ6) a silique was placed in the resin, gently opened and seeds released with the aid of a needle. The pod wall was then discarded. A third cover slip was placed on top so that its edges were supported by the two already on the slide. After the resin was polymerized, the slides were stored at 4°C in the dark, where they remain stable for several months.

Confocal Microscopy: The specimens were imaged with an argon ion laser, 488 nm excitation and 515-530 nm emission, using a Nikon C1 confocal microscope system with a 90i Eclipse microscope and EZ-C1 software (Nikon UK). Images were saved as TIFF files.

2.2.11 Plant DNA extraction

The method was adapted from (Edwards *et al.*, 1991) in order to extract DNA for PCR-based genotyping:

Two to three small, young leaves or an equivalent volume of another tissue, were taken from each plant and placed in a sterile 1.5 ml Eppendorf tube. A toothpick measure of glass beads (Sigma) and 450 µl DNA extraction buffer (200 mM Tris-HCL pH 7.5, 250 mM NaCl and 25 mM EDTA) were added to each tube, and the sample ground using a bench drill and clean plastic pestle. The Eppendorf tube was centrifuged at full speed for 15 min; supernatant was taken, placed in a clean Eppendorf tube and centrifuged at full speed for a further 15 min. 350 µl of the supernatant was transferred to a new Eppendorf tube without disturbing the pellet; an equal volume (350 µl) of absolute isopropanol was then added. The tube was incubated at room temperature for 5 – 30 min. After incubation time, the tube was centrifuged at full speed for 15 min. The pellet was washed three times with 70% ethanol, and then left to dry in a laminar flow hood for 5 – 10 min. 50 – 80 µl of sterile MilliQ water was then added to the pellet to dissolve the DNA. The DNA was either used immediately or stored at 4°C.

2.2.12 Plant RNA extraction

For RNA extraction from plant siliques, fresh siliques were immediately frozen in liquid nitrogen and stored at -80°C until used. The frozen siliques were ground into a powder using sterile ceramic mortar and pestle and RNA was purified using the NucleoSpin® RNA Plant system (Macherey-Nagel, Germany) following the manufacturer's instructions. For RNA extraction from a developing seeds, a silique was opened under a dissecting microscope (Leica MZ6) and the seeds taken out with the aid of a needle and immediately transferred into an Eppendorf tube placed on dry ice, the sample then stored at -80° until use. For more extensive RNA purification, RNA was given a treatment with DNAase to remove any DNA contaminants in the samples by using the RNAase-Free DNAase set (Qiagen) and then concentrated using the RNeasy MinElute Cleanup Kit (Qiagen). The RNA concentration was measured by absorbance at 260 nm using the S2000 UV/Vis Spectrophotometer (WPA) and the purity of the sample ranged from 1.8 – 2.0.

2.2.13 Amplification of DNA for genotyping and expression analyses

PCR was performed using either Taq DNA polymerase (Sigma) or DreamTaq™ DNA polymerase (Fermentas). All reagents were briefly centrifuge after thawing and worked

with on ice. Each 25 µl reaction included 17.75 µl sterile MilliQ water, 2.5 µl of 10X PCR buffer, 0.5 µl of 10 mM dNTP mix, 1 µl of 10 mM forward and reverse primers, 0.25 µl of 5 u/µl Taq DNA polymerase and 2 µl of DNA template. All reactions were placed in the PCR machine (PTC-200 Peltier Thermal cycler, MJ Research). For genotyping purpose a standard amplification programme had an initial denaturation step at 95°C for 5 min, followed by 32 cycles of denaturation step at 94°C for 30 sec, annealing at ~55°C (depending on primer T_m) for 30 sec and extension at 72°C for ~1 min (depending on fragment length, 1 min/kb), and then a final extension at 72°C for 10 min. For expression analyses, the amplification programme consisted of an initial denaturation at 94°C for 3.5 min followed by the same steps as detailed above. Specific PCR conditions are shown in Table 2.2.

2.2.14 cDNA synthesis

cDNA was synthesised from 1 µg of RNA using the ImPromII™ Reverse Transcription System (Promega) following manufacturer's instructions. Then a PCR was performed by using the cDNA as a template (2.2.13).

2.2.15 Agarose gel electrophoresis

1-2% agarose gels were prepared by dissolving agarose (Invitrogen) in 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA, 1.14% glacial acetic acid, pH 7.6); when molten 8 µl Ethidium bromide (10 mg/ml Sigma) was added. The gel was left to set for 1 hour and then placed in a gel tank (Bio-Rad) containing 1X TAE buffer. Samples were loaded with 6X bromophenol blue loading dye. Either a 1 kb or 100 bp DNA ladder (Biolabs) was loaded in the gel as a molecular weight marker. Electrophoresis was carried out at ~100 V for 45 min using a Powerpac 300 power supply (Bio-Rad). DNA bands were visualized using a UV transilluminator (BioDoc).

2.2.16 Genotyping T-DNA and transposon insertion lines

Seeds of T-DNA insertion lines were grown alongside wild type plant as a control. DNA was extracted from the plant leaves, and the genotype verified by PCR. The hypothetical results obtained from the two PCRs that differentiate between wild type, homozygous mutant and heterozygous mutant in gel electrophoresis are as follows. Heterozygous plant present DNA bands in both gels carried both PCRs, whereas a homozygote plant shows

DNA band only in T-DNA PCR product and wild type plant shows band only in genomic PCR product as illustrated in Figure 2.1.

2.2.17 Gel purification of DNA fragments

After separation of DNA fragments by gel electrophoresis, DNA fragments were recovered using the Wizard[®] SV Gel and PCR Clean-Up System (Promega) following the manufacturer's instructions.

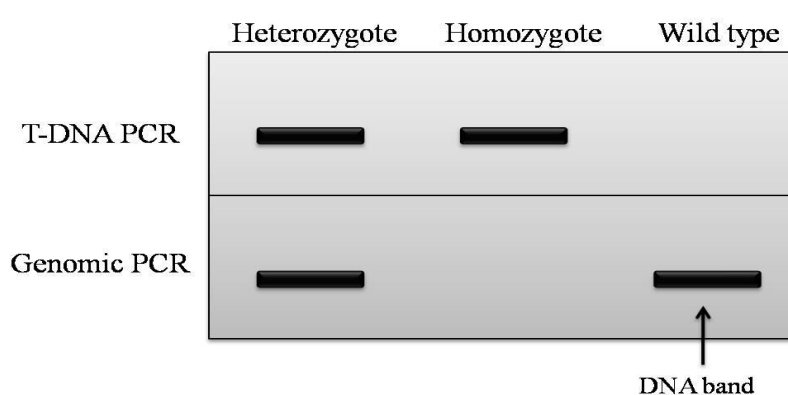


Figure 2.1 Hypothetical diagram of two PCR reaction products loaded in gel. heterozygous sample is represent by bands in both PCRs, homozygous sample is represent a band in the T-DNA PCR only, while wild type sample is represent by a band in genomic PCR only.

Table 2.1 Primer sequence used for genotyping and expression analyses.

Primer name	Sequence (5' – 3')	T _m (°C)
Ds5-1	CCGTTTACCGTTTGTATATCCCG	64.0
CK167	TCCAACACCGAAAACCTCCAT	57.6
CK129	CTAACAAGACAAAGCAAAATACACAAAAG	62.3
LBa1	TGGTTCACGTAGTGGGCCATCG	67.4
SALK_107011_LP	GATTGTGTTCTCTCCTGGTGG	56.6
SALK_107011_RP	GGAACGAGAACTCCATCATTG	56.9
SALK_105945_LP	AAATGTCTGGTTTTATGCCCC	58.5
SALK_105945_RP	GTAGCGAGACAATCGATTTCG	57.8
SALK_116450_LP	TGACAACCTGTAAAATTAGTTGAATACC	60.3
SALK_116450_RP	TAAGCTTCTCCCCTTTTCCAC	58.1
SALK_022148_LP	GATGGTCAAAATGAAAAATGAAAG	58.1
SALK_022148_RP	TTTGGATTGTTTGGTTTGGTTTG	57.9
FLAG_386G01-LP	GCGACTTTTTTCATCAACTTCC	56.8
FLAG_386G01-RP	GAAAATTGTGGATATCCTTCCC	57.3
FST-RB4	TCACGGGTTGGGGTTTCTACAGGAC	69.3
PER-LP	CACAATGGAATGGAGAGAGGAG	58.7
PER/at3g05860-RP	AGGGTTTGAGTTGAACGGAC	56.4
PER/at4g36590-RP	GATTGTGTTCTCTCCTGGTGG	56.6
PER/at5g10440-RP	AAGGGCTTGATTCTGACA	55.1
PER/at2g30810-RP	TGTGAGATGTCGCCGAACA	58.4
PER/at5g46950-RP	GAGTAGTTGCGGCTTTGCTT	57.6
PER/at5g65420-RP	ATCATGCCCTCATCATCAACG	60.0
PER/at1g65300-RP	CTCGGTTAGTTTCTTCGTCATC	56.2
PER/at5g60440-RP	CGACCAGGTGAGAACACAACAT	58.5
PER/at4g02780-RP	CTGAGATGGTAAGGAAAGAAGAAG	56.8
PER/at1g44090-RP	TGCCATTCAAGAACCCACTC	58.1
GAPC-LP	CACTTGAAGGGTGGTGCCAAG	62.7
GAPC-RP	CCTGTTGTGCGCCAACGAAGTC	62.8
pFGC5941-OCS3'/2-Rp	GTTCGCGGCTCGGTGTG	61.1
pFGC5941-OCS-R2	TAGATGTCGCTATAAACCTATTCAG	55.7
AT1G65300_Fp1	TGGAAGGGATCTTCAAGATTTG	58.6
AT4G02780_Fp2	AGGAGATGGGGAAAATGG	54.1
AT5G65420_Fp1	GAGAGAGTGAAGAAGATAGGGGAAATG	62.8
At4g02780-GA1-Fp	GGAGAGAGAGGAGTTGATTGATAAGTG	61.5
AT1G65300_Asc1-F	GGCGCGCCATGAAGAGAAAGATGAAGTTATCGTTA	78.0
65420-QRT-R	TTAGCCTCAAACACAACTG	51.3
65420-QRT-F	ATTTGAATGTTGGAAGAAGAG	51.4
AT4G02780_int-R1	AACCAATACACTTGACTTAGCCC	57.5
35S promoter-FP	CACTATCCTTCGCAAGACCC	57.2
At4g36590-CDS-RP	CCACTTAATGTGCAAAGTTCAC	55.3
At1g65300-CDS-RP	ACCGCACATGCTTCAACAC	57.1
At3g59790-CDS-RP2	AAGAGCGGTTACGGTTC	55.4
At5g46950-CDS-RP2	TTCTTGAAATCATCCTCGC	57.6
At2g19500-CDS-RP	TTGGAATGTACTTTTTCCGTTT	56.5
At3g01330-CDS-RP2	ATACCCGCTTCGGTTCTCC	58.9

Table 2.2 PCR conditions used for genotyping and expression analyses.

PCR description	Primers	Annealing temp (°C)	Extension time (s)	Fragment length (bp)
GAPC expression	GAPC-LP, GAPC-RP	56	40	543
SALK_107011 (T-DNA)	LBa1, SALK_107011_RP	56	30	340
SALK_107011 (genomic)	SALK_107011_LP, SALK_107011_RP	56	40	550
SALK_105945 (T-DNA)	LBa1, SALK_105945_RP	58	60	~840
SALK_105945 (genomic)	SALK_105945_LP, SALK_105945_RP	58	70	~1000
SALK_116450 (T-DNA)	LBa1, SALK_116450_RP	58	60	~800
SALK_116450 (genomic)	SALK_116450_LP, SALK_116450_RP	58	70	~1000
SALK_022148 (T-DNA)	LBa1, SALK_022148_RP	58	60	~800
SALK_022148 (genomic)	SALK_022148_LP, SALK_022148_RP	58	60	1000
FLAG_386G01 (T-DNA)	FLAG_386G01-LP, FST-RB4	57	60	~800
FLAG_386G01 (genomic)	FLAG_386G01-LP, FLAG_386G01-RP	57	70	1070
<i>phe1.1</i> (T-DNA)	Ds5-1, CK167	58	20	180
<i>phe1.1</i> (genomic)	CK129, CK167	58	35	280
<i>35S::MAPK10</i>	35S promoter-FP, At3g59790- CDS-RP2	55.6	60	835
<i>35S::AT5G46950</i>	35S promoter-FP, At5g46950- CDS-RP2	57	60	948
<i>35S::CKX2</i>	35S promoter-FP, At2g19500- CDS-RP	56.5	46	753
<i>35S::E2L2</i>	35S promoter-FP, At3g01330- CDS-RP2	57	50	800
<i>35S::AGL40</i>	35S promoter-FP, At4g36590- CDS-RP	55	45	634
<i>35S::PHE2</i>	35S promoter-FP, At1g65300- CDS-RP	57	45	634
<i>pPER::AGL45</i>	PER-LP, PER/at3g05860-RP	57	60	851
<i>pPER::AGL40</i>	PER-LP, PER/at4g36590-RP	57	60	859
<i>pPER::CYCD4;2</i>	PER-LP, PER/at5g10440-RP	55	60	905
<i>pPER::GA</i>	PER-LP, PER/at2g30810-RP	58	60	870
<i>pPER::AT5G46950</i>	PER-LP, PER/at5g46950-RP	58	60	853
<i>pPER::PHE2</i>	PER-LP, PER/at1g65300-RP	56	50	785
<i>pPER::AGL62</i>	PER-LP, PER/at5g60440-RP	58.5	60	856
<i>pPER::GAI</i>	PER-LP, PER/at4g02780-RP	57	50	783
<i>pPER::CYCD4;1</i>	PER-LP, PER/at5g65420-RP	59	50	654
<i>pPER::GA20OX5</i>	PER-LP, PER/at1g44090-RP	58.5	70	909
<i>PHE2</i> transgene expression	AT1G65300_Fp1, pFGC5941- OCS-R2	57	45	734
<i>PHE2</i> endogenous expression	AT1G65300_Asc1-F, SALK_105945_RP	58	60	781
<i>GAI</i> transgene expression	AT4G02780_Fp2, pFGC5941- OCS-R2	54.5	30	447
<i>GAI</i> endogenous expression	At4g02780-GAI-Fp, AT4G02780_int-R1	58	40	715
<i>CYCD4;1</i> transgene expression	pFGC5941-OCS3'/2-Rp, AT5G65420_Fp1	61	40	586
<i>CYCD4;1</i> endogenous expression	65420-QRT-F , 65420-QRT-R	52	30	260

Chapter 3: Verification of genes involved in endosperm growth and development using a knock-in (KI) strategy

3.1 Introduction

Seed growth is correlated with endosperm growth in crop plants. For example, a major increase in seed volume correlates with the rapid growth of the endosperm (Sundaresan, 2005). Therefore, final seed size could be modified by manipulating endosperm growth. One process that influences final seed size is governed by a parent-of-origin-effect on endosperm development (Scott *et al.*, 1998). In *Arabidopsis thaliana* (*A. thaliana*), interploidy crosses between diploid (2x) and polyploid (4x or 6x) plants produce progeny with an excess of either paternal or maternal genomes, which cause under- or over-proliferation of endosperm nuclei, and either precocious or delayed endosperm cellularisation, resulting in reduced or increased seed size, respectively. Typically, a cross with higher ploidy as the mother results in precocious cellularisation of the endosperm and smaller seeds whereas a cross with higher ploidy as the father exhibits delayed/failed cellularisation and results in larger seeds (Scott *et al.*, 1998).

An important question is which genes contribute to these development phenotypes, and therefore underpin the interploidy cross effect. The availability of microarrays representing almost all the genes expressed from the *A. thaliana* genome enables visualisation of the entire transcriptome of specific tissues and organs. Conducting transcript profiling of seeds manipulated by interploidy crosses thus producing larger or smaller endosperm allows the identification of genes that are potential targets of endosperm manipulation and engineering seed size in crop plants. Previous work in the Scott laboratory identified genes whose transcript abundance either increased, decreased, or remained unchanged in the interploidy crosses at 5 days after pollination relative to the balanced 2x X 2x cross (Tiwari *et al.*, 2010). This chapter focuses on the selection of a subset of genes that display elevated expression levels in large seed crosses (2x X 4x and 2x X 6x) and an initial evaluation of whether these genes play a direct, causal, role in altering endosperm and seed size.

One method with potential for uncovering the role of specific genes is over-expression of the target gene, using constitutive or tissue-specific promoters, in transgenic plants. This knock-in (KI) strategy represents is complementary to knock-out (KO) analysis as a means

to determine gene function (Zhang, 2003). This author suggests that both strategies must be used to realise the great potential of that functional genomics.

Zhang (2003) reviewed several examples of successful KI experiments that had been applied to plants to unravel abiotic stress tolerance, disease resistance, formation of ectopic somatic embryos, ectopic trichome formation and changes in biochemical composition. The, KI strategy used in the work reported here achieved candidate gene over-expression either constitutively from a CaMV 35S promoter or specifically in the young endosperm from a peripheral endosperm specific promoter (*pPER*).

3.1.1 Genes associated with endosperm over-proliferation

Microarray data of *A. thaliana* interploidy crosses was obtained from work carried out at Bath using Affymetrix chips (Tiwari *et al.*, 2010; Section 1.6). Analysis of these data provided a shortlist of candidate genes (Section 3.2.1; Table 3.1). These included *PHERES 2 (PHE2)*, *AGL45*, *AGL40*, *CYCD4;1*, *CYCD4;2*, *GA1*, *GA20-oxidase 5*, *GA-regulated*, *AT5G46950*, *CKX2*, *MAPK10*, and *E2L2* all of which were highly expressed in seeds with a paternalized over-proliferating endosperm and were therefore positively associated with endosperm proliferation (Table 3.2; Figure 3.1). Candidate gene expression was driven either by *pPER* promoter or 35S promoter (Table 3.3). The CaMV 35S promoter is the most frequently used promoter in genetically modified plants. It originates from the cauliflower mosaic virus (CaMV) and is considered a strong constitutive promoter. It leads to high expression of the gene in almost any type of cell and tissue of the plant at any developmental stage (Guilley *et al.*, 1982; Odell *et al.*, 1985; Jefferson *et al.*, 1987; Cornejo *et al.*, 1993; Steinbrecher, 2002). However, despite its wide expression among plant tissues, it may not be expressed in certain tissues, including the endosperm (Benfey *et al.*, 1989; Williamson *et al.*, 1989; Terada and Shimamoto, 1990; Yang and Christou, 1990; Sunilkumar *et al.*, 2002). Hence there was a need to over-express the candidate genes under an endosperm specific promoter. Tiwari *et al.* (2006) identified several promoters that were highly active during the proliferative-phase of endosperm development in *A. thaliana*. Amongst these was the *pPER* promoter isolated from the upstream region of the *AT5G46950* gene that encodes a putative invertase/pectin methylesterase inhibitor protein. The *pPER* drives expression throughout the endosperm (peripheral, chalazal and micropylar regions), but not elsewhere in the plant.

The experimental rationale was as follows: KI/over-expression of growth promoter genes, was expected to produce over-growth phenotypes such as increased leaf size or seed size,

depending on the promoter used for over-expression. In contrast, KI/over-expression of a growth inhibitor gene was expected to have the opposite impact on growth. The potential over-expression lines were accordingly screened for plants displaying promising phenotypes, including large leaf, thick stem and large flower as well as final seed shape and weight compared to a wild type plant and the promising lines were selected for further analysis.

3.2 Results

3.2.1 The selection of genes positively associated with endosperm proliferation

The microarray experiments were discussed earlier in (Chapter1; Section 1.6). From the microarray data generated by hybridising the total RNA from the reciprocal interploidy crosses and the balanced 2x X 2x seeds at 5DAP, ~15,000 genes that were called present in at least one sample above the specified threshold of expression (Tiwari *et al.*, 2010). These genes were subjected to further analysis to identify a manageable shortlist of genes (Table 3.1), with a high probability of direct involvement in the parent-of-origin effect on endosperm growth. The selection criteria were: (1) the genes were associated with endosperm over-proliferation and thus were only up-regulated in paternal excess crosses and did not significantly change or were down-regulated in the maternal excess crosses. (2) the up-regulated genes were confirmed as endosperm-expressed using *in silico* methods such as reported expression in other microarray experiments. (3) the genes had predicted roles in transcriptional activity, cell cycle regulation or hormonal pathway regulation.

Tiwari *et al.* (2010) reported ~114 genes as up-regulated in the 2x X 4x cross which were also not up-regulated in the 4x X 2x and 6x X 2x cross compared to the 2x X 2x cross. We assigned a biological functional category to these 114 genes using the Gene Ontology functionality at TAIR (<http://www.arabidopsis.org/tools/bulk/go/index.jsp>). The categories assigned were as follows; Other cellular processes (40), Other metabolic processes (42), Unknown biological processes (41) Protein metabolism (14), Transcription (13), Cell organization and biogenesis (7), Developmental processes (7), Transport (4), Other biological processes (4), Response to stress (3), Signal transduction (2), Response to abiotic or biotic stimulus (1) where the numbers indicated in brackets are the number of genes in each category. From these candidates were selected that were involved in transcriptional regulation, cell cycle or hormonal regulatory pathway and were also expressed either broadly in the seed or specifically in the endosperm. This resulted in a

shortlist of around 15 genes, namely: *PHERES 1 (PHE1)*, *PHE2*, *AGL45*, *AGL40*, *AGL62*, *AGL28*, *CYCD4;1*, *CYCD4;2*, *GA1*, *GA20-oxidase 5*, *GA-regulated*, *AT5G46950*, *CKX2*, *MAPK10*, and *E2L2*. All the relevant information about the shortlisted candidate genes have been summarised in Tables 3.1 and 3.2, respectively. Following gene selection, the experimental approaches outlined above were used to determine whether the genes had a direct or indirect role in endosperm proliferation.

3.2.1.1 Transcription factor genes

Several genes belonging to MADS box family have been reported to have roles in different aspects of plant developments such as plant reproduction, flowering time, floral organogenesis, floral meristem identity, fruit formation, seed pigmentation and endothelium development (Ng and Yanofsky, 2001; Nesi *et al.*, 2002; Becker and Theissen, 2003). The MADS box genes shortlisted in this study, namely *PHE1*, *PHE2*, *AGL40*, *AGL45*, *AGL62* and *AGL28*, were reported to interact with each other either directly or indirectly (Figure 6.1; de Folter *et al.*, 2005). Similar shared functions of the *AGAMOUS-LIKE* gene clusters have been demonstrated in early endosperm development and have been implicated in interspecific incompatibility, notably *AGL62-AGL80* for endosperm development, *AGL60-AGL90* for interspecific hybridisation and *AGL80* has also been found to interact with *AGL61* to regulate central cell development (Kang *et al.*, 2008; Steffen *et al.*, 2008; Walia *et al.*, 2009).

The selected MADS box genes include *PHE1* and its close homologue *PHE2* (Kohler *et al.*, 2003b), which are both over-expressed in over-proliferative endosperm (Table 3.2). Additionally, *PHE1* is also an imprinted gene, highly expressed paternally and repressed maternally (Kohler *et al.*, 2005). *PHE1* was identified to be over-expressed in *fertilisation independent seed 1/medea (fis1/mea)* mutants that have similar phenotype of paternalized seed in another microarray experiment (Kohler *et al.*, 2003b, Tiwari *et al.*, 2010). *PHE1* and *PHE2* share the same Affymetrix probeset and therefore either transcript may have been responsible for the signal, they are almost 5 fold up-regulated in the 2x X 4x cross (Tiwari *et al.*, 2010). Besides *PHE2*, *AGL40* and *AGL28* were up-regulated in large seeds by nearly 5 folds and their expression was exclusively in endosperm (Table 3.2). The proteins of these genes interacted with *PHE1* and *PHE2* in yeast two-hybrid assays (de Folter *et al.*, 2005). In addition, over-expression of *AGL28* caused precocious flowering (Yoo *et al.*, 2006). *AGL62* is another interesting candidate that has been reported to be 1.6 times up-regulated in over-proliferative endosperm in our microarray data (Tiwari *et al.*,

2010), its expression has been found to be exclusively in the endosperm as reported by several microarray studies (Table 3.2). AGL62 protein interacts directly with PHE1 and PHE2 proteins (de Folter *et al.*, 2005). Loss-of-function of AGL62 has been reported to cause precocious endosperm cellularisation resulting in seed abortion. This has been independently verified in parallel in our study outlined in Chapter 5 (Kang *et al.*, 2008). Further to these AGL45, which encodes a protein reported to interact with AGL40 and is expressed exclusively in endosperm tissues (Table 3.2) was also found to be 5 fold up-regulated in the big seeds.

Examples of other transcription factors that belong to the same family (MADS box, type 1) such as AGL80 (Portereiko *et al.*, 2006) and AGL23 (Colombo *et al.*, 2008) provided good evidence of MADS box gene in controlling gametophyte or/and early seed development (discussed in Chapter 6). Therefore, it is important to study the functions of the other genes grouped within this family either as single genes or within a complex. Further experiments such as study of mutants of one or more members of the complex are necessary to examine whether up-regulation of the selected MADS box genes in interploidy crosses is a cause or a consequence of enhanced seed growth.

3.2.1.2 Genes involved in cell cycle

Cell cycle development is controlled by reversible phosphorylation by protein kinases and phosphatases, and particularly by the serine-threonine kinase activity of cyclin dependent kinase (CDK) complexes. CDK complexes contain a catalytic CDK subunit, and a regulatory cyclin subunit. Their activity is further modified by phosphorylation and the binding of inhibitory and scaffolding proteins (Morgan, 1997; Menges *et al.*, 2005).

Of the 114 genes which were up-regulated in paternalized endosperm there were genes involved in cell cycle processes such as cell proliferation. Although paternalized endosperm showed increased and prolonged proliferation, and delayed or inhibited cytokinesis, few core cell cycle regulators were over-expressed in over-proliferative endosperm (Tiwari *et al.*, 2010). The core cell cycle regulators up-regulated in the big seeds included two D-type cyclins, *CYCD4;1* and *CYCD4;2*. The D-type cyclins control progression from the early G1 phase to S phase during mitotic cycle response to external growth signals (cytokinin and sucrose) (Lara-Nunez *et al.*, 2008). *CYCD4;1* was reported to be rate-limiting for cell division in root apex of germinating seeds and gene loss of function showed a delay in cell proliferation as well as a reduction in total cell number

(Masubelele *et al.*, 2005). Recently, *CYCD4;1* expression was also reported in meristematic pericycle protoxylem poles and is required for normal lateral root (LR) density by normal spacing but not initiation (Nieuwland *et al.*, 2009). The authors also found that *CYCD4;1* expression is sucrose-dependent and wild type roots showed *CYCD4;1* loss-of-function phenotype in sucrose absence. The authors concluded that *CYCD4;1* associate meristem pericycle cell behaviour to LR density consistent with a basal meristem pre-patterning model and that D-type cyclins can confer division potential of defined cell types through cell-specific expression patterns.

In addition, *CYCD4;2* has previously been reported to have a unique protein structure which lacks the retinoblastoma (Rb)-binding motif and PEST sequence that are hallmarks of other D cyclins (Kono *et al.*, 2006). However, functional assays of KI and KO plants showed that *CYCD4;1* had a role in proliferation as a cyclin regardless its unique structure. (Kono *et al.*, 2006; Kono *et al.*, 2007). The *CYCD4;1* and *CYCD4;2* expression were reported to be absent or at low level in other microarray experiments (www.genevestigator), while in this work both genes showed a strong association with seed growth which make them an interesting genes in controlling cell proliferation in developing seeds (Table 3.2). *CYCD4;1* was found to be 1.5 fold more whereas *CYCD4;2* was more than 2 fold up-regulated in the 2x X 4x (paternalized) cross.

E2L2 (DEL3) is one of the interesting genes in our list that is involved in cell cycle processes and found to be 1.5 fold up in large seeds and down-regulated in small seeds (Tiwari *et al.*, 2010; Table 3.2). *E2L2* is belongs to E2F/dimerization partner (DP) transcription factors family which control cell cycle transitions in both animals and plants either negatively or positively (Ramirez-Parra *et al.*, 2004). Modulation of E2F/DP activity was originated by either Rb-repressors, which were responsible in blocking the activation domain of E2Fs and/or by recruiting chromatin remodelling factors (Harbour and Dean, 2000; Rossi *et al.*, 2003). The *A. thaliana* E2F/DP family consists of three AtE2F genes namely: AtE2Fa/AtE2F3, AtE2Fb/AtE2F1 and AtE2Fc/AtE2F2, two DP genes and three E2F-like genes known as DELs, E2Ls, or E2Fd-f (Wellmer and Riechmann, 2005). Typical E2F proteins heterodimerize with DP proteins to bind E2F sites in promoters of genes associated with DNA synthesis and replication and cell cycle control including D cyclins, and may be positive or negative regulators of cell division (Kosugi and Ohashi, 2002; Mariconti *et al.*, 2002). In contrast, atypical E2Fs proteins such as DEL or E2F was reported to have a negative activity in the E2F network by causing

gene repression in which they can bind to DNA in a monomeric form under E2F control (Kosugi and Ohashi, 2002).

In a protein fluorescence assay, E2L2 was detected in both the nucleus and cytoplasm (Kosugi and Ohashi, 2002). The author suggested that E2Ls may balance the activities of E2F-DP and play a role in limiting cell proliferation because the transcripts of E2Ls were abundant in meristematic rather than fully differentiated tissues (Table 3.2). The function of DEL3/E2Ff has not yet been described.

MAPK10 gene encodes a mitogen-activated protein kinase (MAPK) and its transcripts are detected in silique, endosperm, seed coat and stem as reported by different microarray experiments (www.genevestigator). In addition, *MAPK10* expression was up-regulated more than 4 folds in the 2x X 4x cross and almost 2 fold in the 2x X 6x but not in maternalized crosses (Tiwari *et al.*, 2010; Table 3.2). MAPK signal transduction cascades are known regulators of various aspects of plant biology reviewed by Andreasson and Ellis (2010). In addition, these protein phosphorylation cascades link extracellular stimuli to a wide range of cellular responses (Feilner *et al.*, 2005). The function of *MAPK10* has not yet been described.

3.2.1.3 Genes involved in cell signalling or hormonal pathways

The other group of candidate genes in this study were involved in cell signalling or hormone pathways. Tiwari *et al.* (2010) reported that the cytokinin oxidase *CKX2* was more than 1.5 times up-regulated in paternalized endosperm and down-regulated or remained unchanged in maternalized endosperm (Table 3.2). *CKX2* encodes a cytokinin oxidase/dehydrogenase protein which regulates cell division and proliferation by affecting cytokinin hormone activity (Werner *et al.*, 2003). Recently, Werner *et al.* (2006) demonstrated that CKX overexpression increases seed size in *A. thaliana* to a greater extent than can be completely accounted for by the accompanying loss of fertility.

Other interesting genes were involved in gibberellic acid (GA) metabolism. Gibberellic acid is an essential phytohormone which controls many aspects of plant development, including seed germination, leaf expansion, stem elongation, flowering, and seed development (Ogawa *et al.*, 2003). The genes that were up-regulated in large seeds included: *GA20OX5* (3.5 fold up) encoding a member of the GA 20-oxidase family which catalyses synthesis of bioactive GA, *GAI* (1.2 fold up) encoding a copalyl diphosphate synthase that catalyses the first committed step in GA biosynthesis, and *GA-regulated*

encoding a GA-regulated family protein (more than 5 fold up). In other microarray experiments, *GA20OX5* was also expressed exclusively in endosperm, while *GAI* was expressed in endosperm and stem (Table 3.2). The *GA-regulated* gene had a particularly strong association with seed growth, showing very high levels of over-expression in microarray data, 6-fold in 2x X 4x, 20-fold in 2x X 6x, and severe under-expression in maternalized endosperm (Tiwari *et al.*, 2010; Table 3.2). Other microarray experiments show this gene is highly expressed in siliques and also present in isolated seeds (www.genevestigator) but no function has been reported.

3.2.1.4 Genes involved in other biological process

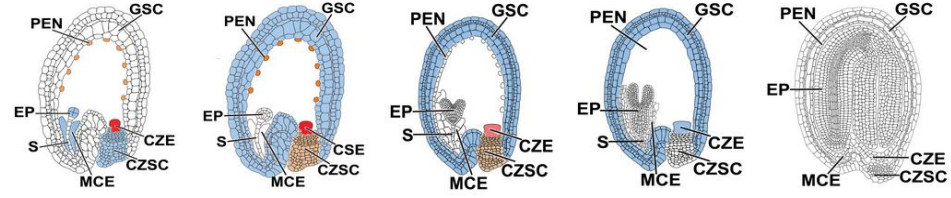
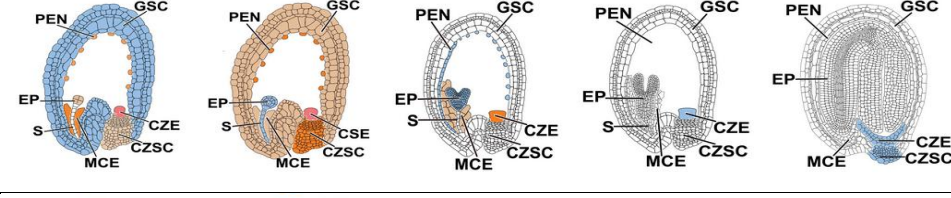
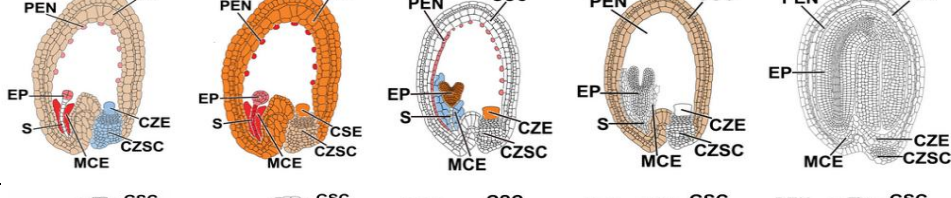
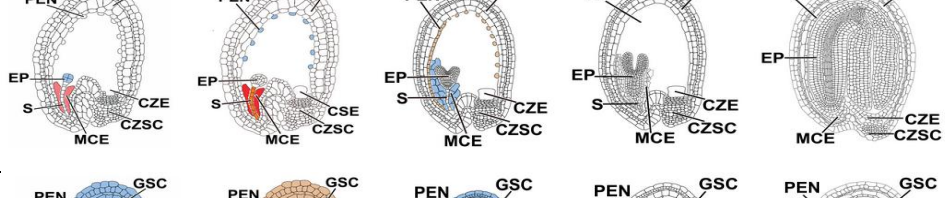
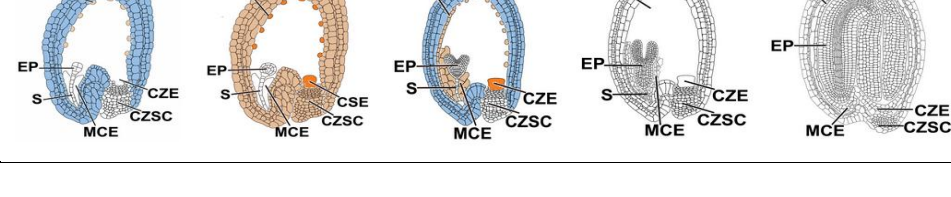
An interesting candidate gene that showed high expression in large seed and not in small seed was At5g46950, which encodes an invertase/pectin methylesterase inhibitor family protein. This gene is expressed exclusively in endosperm (Table 3.2; Tiwari *et al* 2006). In general, pectin methylesterases (PMEs) are ubiquitous enzymes that alter the degree of methylesterification of pectins, which are major components of plant cell walls (Micheli, 2001; Pelloux *et al.*, 2007). Pelloux *et al.* (2007) also added that PMEs are implicated, directly and indirectly, in different physiological processes associated with both vegetative and reproductive plant development. The function of At5g46950 has not yet been described but was one of the genes whose upstream region was used as the promoter (*pPER*) to drive the expression of the other candidate genes.

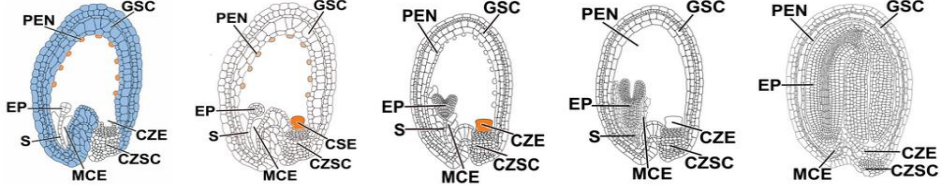
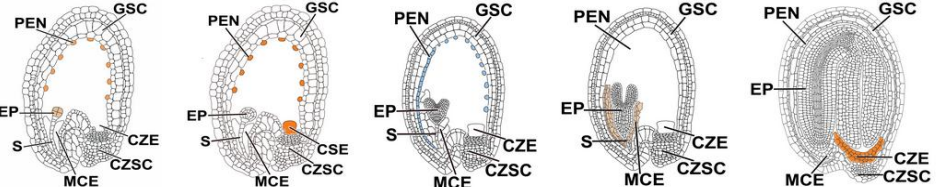
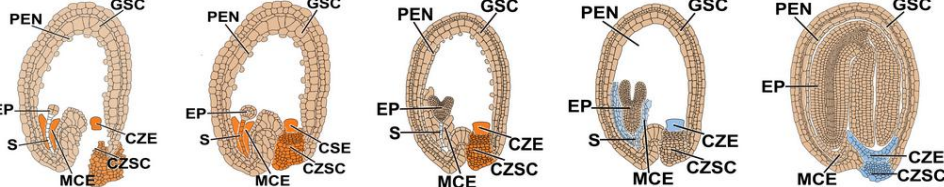
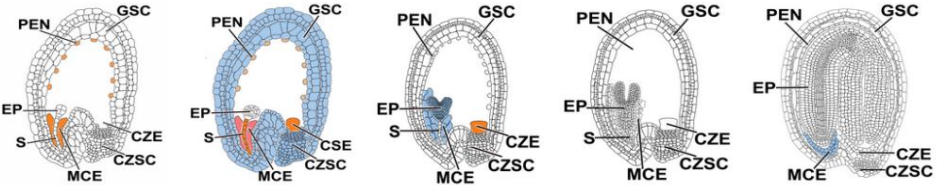
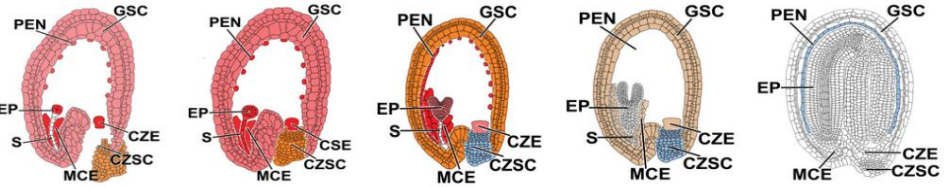
Table 3.1 General information of the selected candidate genes.

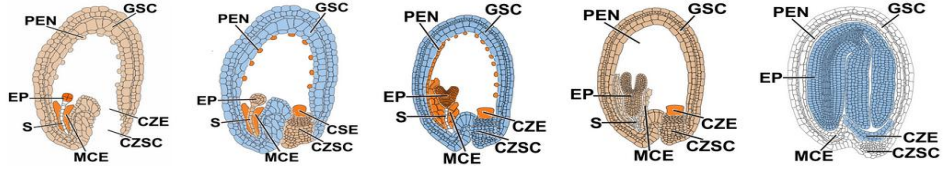
AtId	Symbol	Description	Comments	References
At1g65330	<i>PHE1/AGL37</i>	Type 1 MADS box gene involved in DNA binding /other name <i>AGL37</i>	Imprinted gene, regulated by MEA and FIE, expressed transiently after fertilization in embryo and endosperm	Kohler <i>et al.</i> , 2003; 2005
At1g65300	<i>PHE2/AGL38</i>	Type 1 MADS box gene involved in DNA binding/ other name <i>AGL38</i>	Transcription factor activity	
At5g60440	<i>AGL62</i>	Type 1 MADS box gene involved in DNA binding	Controls cellularisation during the syncytial phase of endosperm development	Kang <i>et al.</i> , 2008
At1g01530	<i>AGL28</i>	Type 1 MADS box gene involved in DNA binding	Overexpression of <i>AGL28</i> caused precocious flowering	Yoo <i>et al.</i> , 2006
At3g05860	<i>AGL45</i>	Type 1 MADS box gene involved in DNA binding	Transcription factor activity	
At4g36590	<i>AGL40</i>	Type 1 MADS box gene involved in DNA binding	Transcription factor activity	
At5g46950		Invertase/pectin methylesterase inhibitor family protein	Unknown function	
At2g19500	<i>CKX2</i>	Cytokinin oxidase family protein	Cytokinin catabolic process	Werner <i>et al.</i> , 2003
At3g59790	<i>MAPK10</i>	Mitogen-activated protein kinase	Unknown function	
At1g44090	<i>GA20OX5</i>	Gibberellin 20-oxidase family protein	GA20ox enzyme catalyses a series of intermediate oxidation reactions during the biosynthesis of the phytohormone gibberellin	
At4g02780	<i>GAI</i>	Gibberellin-requiring 1; ent-copalyl diphosphate synthase	Involved in the first committed step in GA biosynthesis	
At5g65420	<i>CYCD4;1</i>	Cyclin-independent protein kinase regulator	Involved in G1 phase of mitotic cell cycle, regulation of cell cycle, response to sucrose stimulus and stomatal lineage progression	Masubelele <i>et al.</i> , 2005; Nieuwland <i>et al.</i> , 2009
At5g10440	<i>CYCD4;2</i>	Cyclin-independent protein kinase regulator	Encodes a cyclin involved in cell proliferation during stomatal cell lineage development.	Kono <i>et al.</i> , 2006; 2007
At2g30810	<i>GA-regulated</i>	Gibberellin-regulated family protein	Involved in response to gibberellin stimulus	
At3g01330	<i>E2L2</i>	Putative/E2F-like repressor, other name <i>DEL3</i>	Member of the E2F transcription factors, (cell cycle genes), key components of the cyclin D/retinoblastoma/E2F pathway	Ramirez-Parra <i>et al.</i> , 2004

Table 3.2 *In silico* expression analysis of the selected candidate genes.

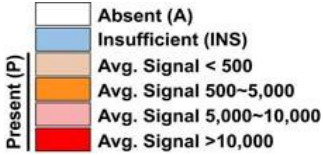
AtId	Symbol	Interploidy crosses ^a				In various stage/tissue compartment of seed ^b					Other web recourses ^c
		6xX2x	4xX2x	2xX4x	2xX6x	Pre-globular stage	Globular stage	Heart stage	Linear cotyledon stage	Maturation green stage	
At1g65330 *	<i>PHE1</i>	-4.2	-4.6	2.3	2.4						Endosperm
At1g65300 *	<i>PHE2</i>	-4.2	-4.6	2.3	2.4						Endosperm
At5g60440	<i>AGL62</i>	1.1	-0.2	0.7	2.8						Endosperm
At1g01530	<i>AGL28</i>	-0.1	-0.9	2.3	0.7						Endosperm

At3g05860	<i>AGL45</i>	-0.9	-0.8	2.3	2.7		Endosperm
At4g36590	<i>AGL40</i>	-2.8	-3.0	2.4	3.6		Endosperm
At5g46950		-1.2	-3.3	2.3	2.7		Endosperm, suspensor
At2g19500	<i>CKX2</i>	0.0	-1.2	0.8	2.2		Sperm cell, endosperm, suspensor,
At3g59790	<i>MAPK10</i>	-1.7	-2.7	2.1	0.7		Sperm cell, stamen, endosperm

At1g44090	GA20OX5	-0.7	-1.2	1.8	3.3		Endosperm, root endodermis
At4g02780	GAI	-0.1	0.3	0.3	2.3		Endosperm
At5g65420	CYCD4;1	0.1	-0.7	0.6	1.2		Sperm cell, seedlings, flower, seed, leaf, stem, root
At5g10440	CYCD4;2	0.1	-0.9	1.1	1.4		Endosperm, suspensor
At2g30810	GA-regulated	-3.1	-4.3	2.5	4.3		Endosperm, embryo, seed coat, suspensor

At3g01330	E2L2	-3.1	-1.1	0.6	0.5		Sperm cell, inflorescence, embryo, endosperm, seed coat, shoot apex, root tip
-----------	------	------	------	-----	-----	--	---

* The probe set cannot differentiate between these genes in analysis. ^a Single log2 ratio (SLR) of expression compared to 2x X 2x cross in 5DAP siliques (Tiwari *et al.*, 2010). ^b Gene Chip experiments done in Goldberg and Harada laboratories (www.seedgenenetwork.net). ^c Expression profile in other tissue/organ in plant rather than seed available in (GENENVISTEGATOR and Arabidopsis eFP Browser). Tissue: CZE - Chalazal Endosperm; CZSC - Chalazal Seed Coat; EP - Embryo Proper; GSC - General Seed Coat; MCE - Micropylar Endosperm; PEN - Peripheral Endosperm; S - Suspensor; WS - Whole Seed.



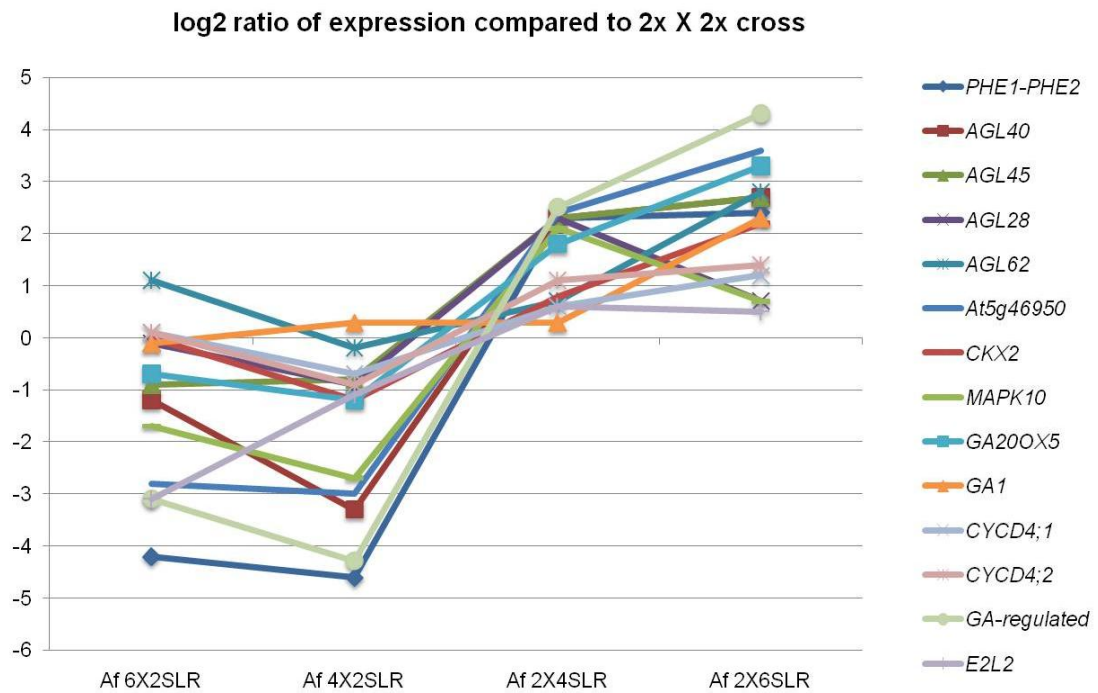


Figure 3.1 Trends in expression of selected genes in the paternal and maternal excess crosses. (Data courtesy Tiwari *et al.*, 2010).

Table 3.3 Candidate genes and their KO and KI insertion lines.

AtId	Symbol	KO	KI
At1g65330	<i>PHE1</i>	Ds5 transposon	Not available ^a
At1g65300	<i>PHE2</i>	T-DNA	35S promoter ^b <i>pPER</i> ^c
At4g36590	<i>AGL40</i>	T-DNA	35S promoter <i>pPER</i>
At3g05860	<i>AGL45</i>	T-DNA	<i>pPER</i>
At1g01530	<i>AGL28</i>	T-DNA	Not available
At5g60440	<i>AGL62</i>	T-DNA	<i>pPER</i>
At5g46950		Not available	35S promoter <i>pPER</i>
At2g19500	<i>CKX2</i>	Not available	35S promoter
At3g01330	<i>E2L2</i>	Not available	35S promoter
At3g59790	<i>MAPK10</i>	Not available	35S promoter
At1g44090	<i>GA20OX5</i>	Not available	<i>pPER</i>
At4g02780	<i>GA1</i>	Not available	<i>pPER</i>
At5g65420	<i>CYCD4;1</i>	Not available	<i>pPER</i>
At5g10440	<i>CYCD4;2</i>	Not available	<i>pPER</i>
At2g30810	<i>GA-regulated</i>	Not available	<i>pPER</i>

^a Not available means not included in this study.

^b Provided as T2 and T3 from Ceres Inc., USA.

^c Provided as T0 and T1 from Dr. Sushma Tiwari, University of Bath, UK.

3.2.2 Phenotypic analysis of plants containing 35S promoter driven candidate genes

The KI strategy was designed to result in over-expression of the target gene in wider tissues by using the 35S promoter. Where the target gene is a growth promoter the expectation was that over-expression would result in cell/organ over-proliferation. In contrast, where the KI is for growth inhibitor gene, the gene over-expression will produce small vegetative and reproductive organ. The *A. thaliana* lines that were used as the 35S over-expression insertion lines for the genes (*PHE2*, *AGL40*, *AT5G46950*, *CKX2*, *MAPK10* and *E2L2*) were derived from Col-0 ecotype and obtained from Ceres Inc, California, USA as T2 segregating seeds (Table 3.4). They have the *bar* gene conferring resistance to the herbicide BASTA as a selection marker.

Seeds of the individual independent insertion lines (illustrated in Table 3.4) were planted alongside the wild type and subjected to BASTA selection to isolate the resistant plants containing the respective transgenes for further analysis. Only over-expression independent lines that showed a different phenotype than wild type such as leaf shape, leaf size, stem length, flower size, etc were selected for additional analysis. The presence of the transgenes in the individual independent lines was verified by PCR using specific primers (Figure 3.2) (the primers and PCR conditions are detailed in Table 2.1 and Table 2.2, respectively). All the BASTA resistant plants recovered also tested positive for the presence of the transgene when verified by PCR.

KI lines of *PHE2*, *AGL40* and *AT5G46950*, *E2L2* (ME10408-01 and -02), *CKX2* (ME11512-01 to 04) and *MAPK10* (ME15230-01 and 03-06) yielded plants that were indistinguishable from wild type under normal growth conditions (Table 3.4). In contrast, the over-expression lines of *E2L2*, *CKX2* and *MAPK10* showed abnormal phenotype and are discussed in the next section.

3.2.2.1 35S-KI plants of *E2L2*, *MAPK10* and *CKX2* resulted in vegetative abnormalities

Plants with obvious abnormal phenotypes were found among the KI lines of the *E2L2*, *MAPK10* and *CKX2* genes. The morphology of several features of these contrasted with the wild type Col-0 plant is shown in Figure 3.3.

The *E2L2* (ME10408-03) KI line developed progeny with abnormal phenotypes at a frequency of 22% compared to wild type (Col-0) (Table 3.4). Figure 3.4, shows an *E2L2* plant with a curly woody stem and apparently sterile flowers. However, when the plant was

pollinated by wild type pollen normal numbers of seeds developed in the siliques (Figure 3.7b).

Among the *MAPK10* KI lines, ME15230-02 produced T2 progeny with abnormal aboveground phenotypes (Table 3.4). The ME15230-02 plants produced longer stems, long pistils and very short siliques compared to wild type plants (Figure 3.5). The dry silique of the *MAPK10*-ME15230-02 plants contained relatively few but large seed indicating reduced fertility (Figure 3.7d).

About 50% of *CKX2*-ME11512-05 plants were highly stunted, with 21 day old plants typically consisting of a few pairs of small leaves; flowering was early by approximately 5 days and flowers were small with relatively long pistils (Figure 3.6). These plants rarely produced seeds by self-pollination, but when assisted in pollination produced numerous highly abnormal and shrivelled seeds (Figure 3.7d).

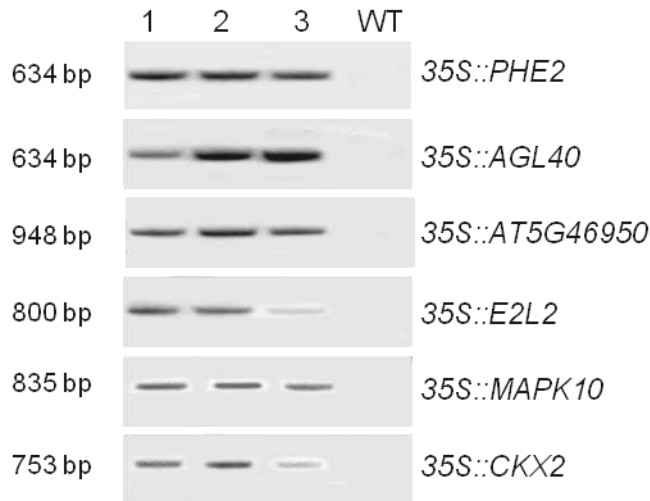


Figure 3.2 PCR products of KI lines driven by the 35S promoter. Examples from *35S::PHE2*, *35S::AGL40*, *35S::AT5G46950*, *35S::E2L2*, *35S::MAPK10* and *35S::CKX2* plants are shown. Each band represents a positive transformation in different T2 plants. WT is wild type Col-0, 1kb DNA ladder was used (band size on the left). For DNA amplification, 35S promoter-FP primer and the reverse specific primer for each transgene were used (the primers and PCR conditions are detailed in Table 2.1 and Table 2.2, respectively).

Table 3.4 Summary of genes driven by 35S promoter.

Construct	Ceres line ID	Individual independent line ^a	Generation	The proportion of normal to abnormal phenotype ^b
<i>35S::AGL40</i>	ME04265	ME04265-02	T2	All normal
		ME04265-03	T2	All normal
		ME04265-10	T2	All normal
<i>35S::PHE2</i>	ME15235	ME15235-01	T2	All normal
		ME15235-02	T2	All normal
		ME15235-03	T2	All normal
		ME15235-04	T2	All normal
		ME15235-05	T2	All normal
		ME15235-06	T2	All normal
		ME15235-07	T2	All normal
		ME15235-08	T2	All normal
<i>35S::AT5G46950</i>	ME12746	ME12746-01	T2	All normal
		ME12746-02	T2	All normal
		ME12746-03	T2	All normal
		ME12746-04	T2	All normal
		ME12746-05	T2	All normal
	ME12770	ME12770-01	T2	All normal
		ME12770-02	T2	All normal
		ME12770-03	T2	All normal
		ME12770-04	T2	All normal
		ME12770-05	T2	All normal
	ME04363	ME04363-01	T2	All normal
		ME04363-02	T2	All normal
		ME04363-03	T2	All normal
	ME12793	ME12793-01	T2	All normal
		ME12793-02	T2	All normal
		ME12793-03	T2	All normal
		ME12793-04	T2	All normal
		ME12793-05	T2	All normal
<i>35S::E2L2</i>	ME10408	ME10408-01	T2	All normal
		ME10408-02	T2	All normal
		ME10408-03	T2	78% normal:22% abnormal
<i>35S::CKX2</i>	ME11512	ME11512-01	T2	All normal
		ME11512-02	T2	All normal
		ME11512-03	T2	All normal
		ME11512-04	T2	All normal
		ME11512-05	T2	54% normal:46% abnormal
<i>35S::MAPK10</i>	ME15230	ME15230-01	T2	All normal
		ME15230-02	T2	All abnormal
		ME15230-03	T2	All normal
		ME15230-04	T2	All normal
		ME15230-05	T2	All normal
		ME15230-06	T2	All normal

^a Individual independent lines were not single insertion homozygous lines.

^b Overall plant phenotype such as stem length, leaf shape and size, number of branches, flower symmetry, and flowering time compared to wild type (Col-0; Figure 3.3); normal = wild type phenotype, abnormal = different than wild type phenotype.

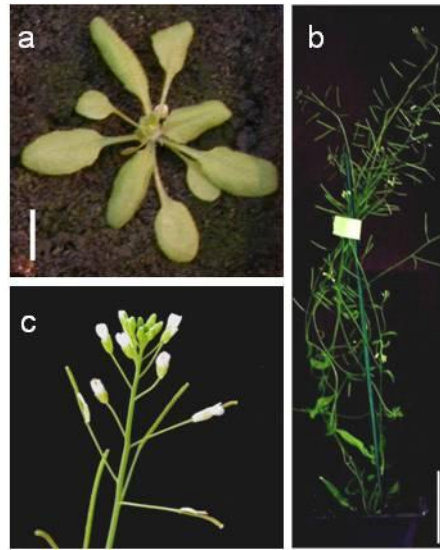


Figure 3.3 Morphology of the wild-type (Col-0) plants. (a) A 21 day-old plant. (b) The overall morphology of a 49 day-old plant showing well developed inflorescences bearing numerous well elongated siliques. (c) The inflorescence. Bar in a = 1 cm and in b = 4 cm.

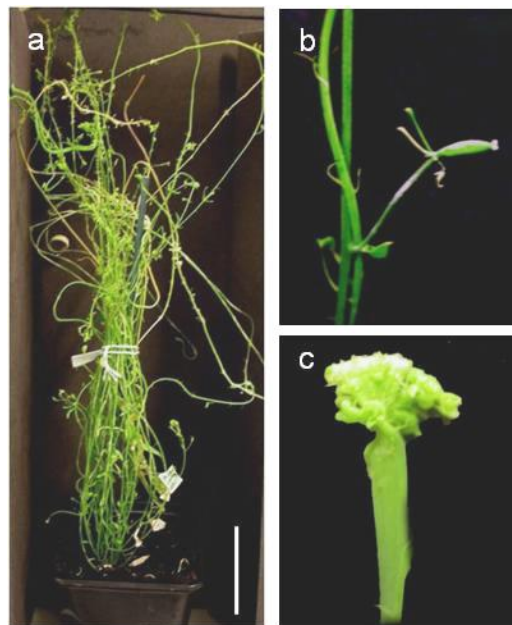


Figure 3.4 Morphology of the plants from *E2L2* (ME10408-03) KI line. (a) A 49 day-old plant showing a woody, curly stem, small leaves and abnormal flowers. Bar = 5 cm. (b) Flower consisting of a pistil only. (c) A highly aberrant inflorescence.

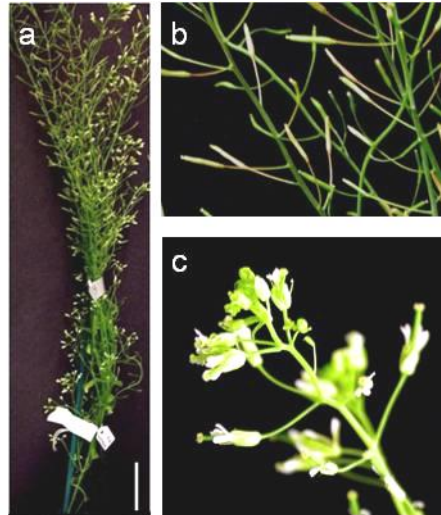


Figure 3.5 Morphology of plants from the *MAPK10* (ME15230-02) KI line. (a) 41 day-old plant. Bar size = 4 cm. (b) Short siliques. (c) An inflorescence showing flowers containing long pistils.

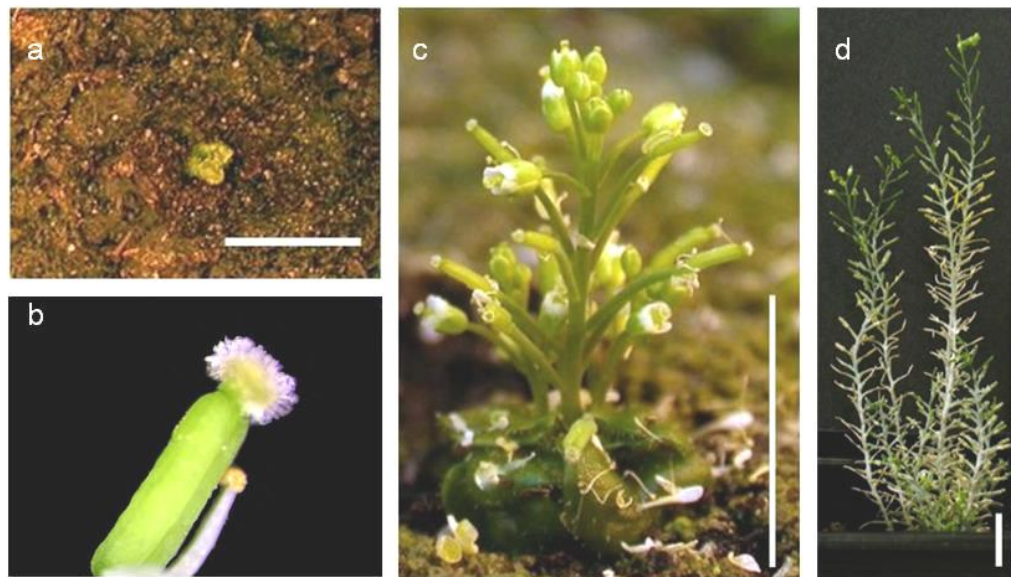


Figure 3.6 Morphological features of *CKX2* (ME11512-05) KI plants. (a) A 21 day-old plant displaying a highly stunted phenotype. (b) Flower showing the relative over-growth of the pistil compared to the stamen. (c) 37 day-old plant showing the persistent and severe stunting associated with this line (d) A 65 day-old is 25% the height of wild type Col-0 (see Figure 3.3c). Bar in a, c and d = 1 cm.

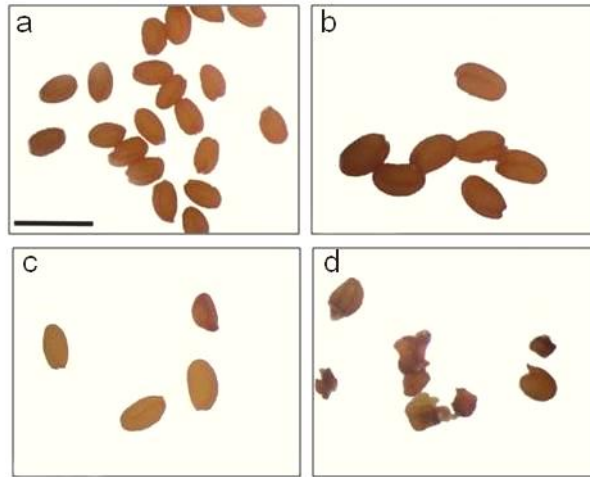


Figure 3.7 Mature seeds of wild type and 35S KI lines. (a) Wild type (Col-0) seeds, (b) Seeds from a *E2L2* X Col wild type cross (c) *MAPK10* selfed seeds, (d) *CKX2* selfed seed. Bar = 1mm.

3.2.3 The *pPER* promoter

The aim of using the *pPER* endosperm-specific promoter was to over-express the candidate genes in the endosperm to test their role in endosperm development. The upstream region of At5g46950 was fused with the cDNA of the candidate genes and assembled in a binary vector. The vector also contained the *bar* gene which confers resistance to the herbicide BASTA and thus acts as a selectable marker for transformation. Independent primary transformants were selected using BASTA and used for further screening. A previous study conducted in the Scott laboratory (Robert Jackson and Kate Breed, personal communication) screened individual independent lines for the candidate genes. Individual independent T1 lines that showed significantly heavier seeds than wild type were selected for further analysis namely: *pPER::PHE2*, *pPER::CYCD4;1*, *pPER::GA1* and *pPER::GA20OX5* lines. The T2 and T3 plants of the selected over-expression lines were grown along with the wild type and then sprayed with BASTA to identify homozygous single insertion lines within the T3 generation (100% resistance to BASTA) (Chapter 2; Section 2.2.8). The plants were then allowed to self pollinate, and subjected to seed phenotype analysis.

3.2.3.1 Analysis of *pPER::PHE2* plants

3.2.3.1.1 Seed size analysis

To determine whether over-expression of *PHE2* in the endosperm altered seed size, four independent T3 *pPER::PHE2* lines that were homozygous for a single insertion and showed promising high seed weight in the previous generation were analysed for increased seed weight. All lines produced seed that was either significantly heavier or lighter than wild type (Figure 3.8a). *pPER::PHE2*-A9/P2/P1, *pPER::PHE2*-A9/P1/P3 and *pPER::PHE2*-A7/P6/P6 produced heavier seeds compared to wild type Col-0 ($p < 0.001$, 0.001 and 0.002, respectively). Whereas *pPER::PHE2*-A8/P5/P3 produced lighter seeds than wild type ($p < 0.001$).

A trade-off exists between seed number and seed size, where a reduction in seed number results in an increase in seed weight, and *vice versa* (Venable, 1992). To test whether the observed seed size phenotypes resulted from a direct effect of transgene expression or from an indirect effect on seed number, the seed number per silique was measured for each over-expression line and the wild type control (Figure 3.8b). The data showed that the fertility of *pPER::PHE2*-A8/P5/P3 and *pPER::PHE2*-A7/P6/P6 was not compromised as more than 70 seeds were produced per silique, which represented no significant difference to wild type ($p = 0.118$ and 0.410, respectively). Consequently, the increase in seed weight of *pPER::PHE2*-A7/P6/P6 line over wild type observed was not a result of reduced seed number.

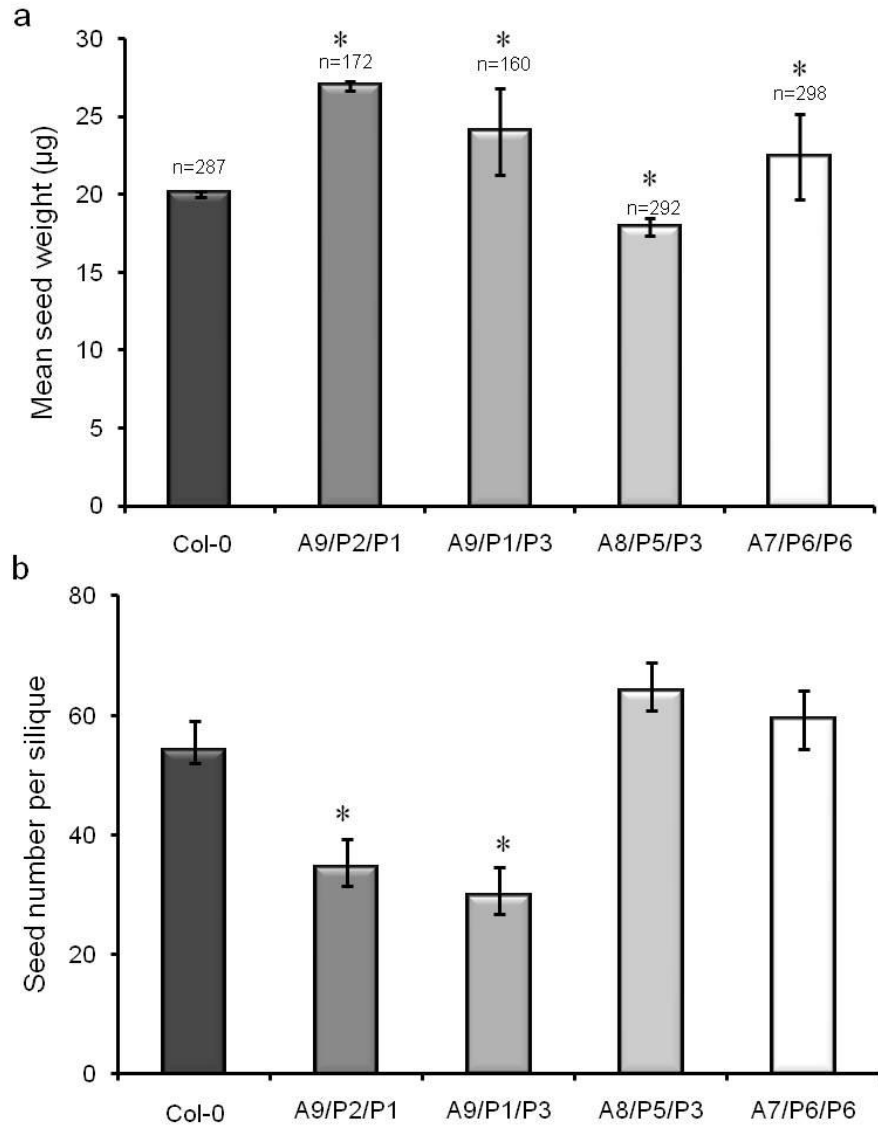


Figure 3.8 Seed weight analysis of four independent, homozygous, single insertion *pPER::PHE2* plants. (a) Mean seed weight; n = number of seeds examined. (b) Seed set per silique (n = 5 siliques of each of 3 plants). For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *. Error bar = S.E.M.

In contrast, *pPER::PHE2-A9/P2/P1* and *pPER::PHE2-A9/P1/P3* were less fertile compared to wild type Col-0, with an average reduction in seed number per silique of 37% ($p < 0.001$) and 44% ($p < 0.001$), respectively. Therefore, the increase in seed weight over wild type could have resulted from reduced seed number.

To determine whether the changes in seed size observed in *pPER::PHE2-A7/P6/P6* and *pPER::PHE2-A9/P2/P1* were correlated with an increase in endosperm proliferation, the size of the peripheral endosperm was estimated using the embryo sac area of 6 DAP seeds

as a proxy measure (Scott *et al.*, 1998). The embryo sac area of seeds developing on *pPER::PHE2-A9/P2/P1* plants was significantly larger than wild type seeds at 6 DAP (13% increase; $p < 0.001$) (Figure 3.9a). In contrast, the embryo sac area of *pPER::PHE2-A7/P6/P6* developing seeds was similar to wild type (Col-0; $p = 0.905$) (Figure 3.9b). One explanation for the inconsistency of the correlation between seed weight and embryo sac area could be that the embryo sac area measures the endosperm size only but the final seed weight includes contributions from the embryo, endosperm and the seed coat (integument layer).

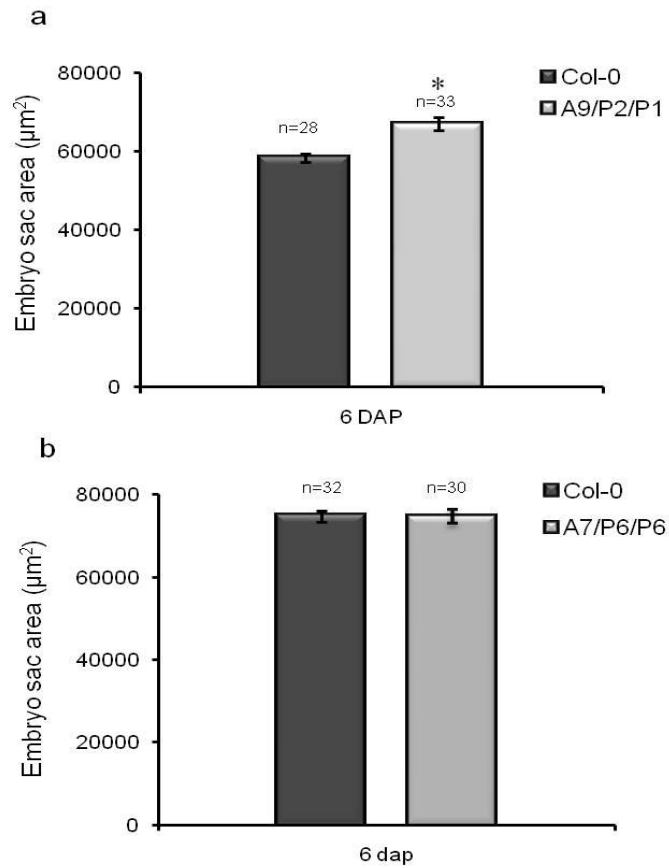


Figure 3.9 Embryo sac area of wild type and two independent, homozygous single insertion *pPER::PHE2* plants at 6 DAP. (a) Mean embryo sac area of *pPER::PHE2-A9/P2/P1* and Col-0. (b) Mean embryo sac area of *pPER::PHE2-A7/P6/P6* and Col-0. For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *; n = number of seeds examined. Error bar = S.E.M.

3.2.3.1.2 Expression of *PHE2* in *pPER::PHE2* plants

To determine whether the *pPER::PHE2* transgene resulted in an increase in *PHE2* mRNA levels within the endosperm of transgenic plants, semi-quantitative RT-PCR was carried out on cDNA synthesized from RNA extracted from 5 DAP siliques. Three lines that were

positive for the presence of the transgene, *pPER::PHE2*-A9/P2/P1, *pPER::PHE2*-A9/P1/P3 and *pPER::PHE2*-A7/P6/P6, were tested. The level of the transgene expression was determined using two primers (AT1G65300_Fp1 and pFGC5941-OCS-R2) to amplify a sequence between the gene and the OCS 3' terminator, (Figure 3.10a) (the primers and PCR conditions are detailed in Table 2.1 and Table 2.2, respectively). Since one of the primers targeted the OCS, this primer pair could not amplify cDNA derived from the endogenous *PHE2* gene. The overall *PHE2* expression level, endogenous and transgene-driven expression, was determined using two primers (AT1G65300_Asc1-F and SALK_105945_RP) that anneal and amplify a sequence within the gene coding region (Figure 3.10b). Three independent transformation lines were analysed. Whilst all tested positive for transformation, only *pPER::PHE2*-A7/P6/P6 showed detectable transgene expression (Figure 3.10c). Only the A7/P6/P6 line showed over-expression of *PHE2* in the semi-quantitative RT-PCR, providing strong supporting evidence that the *pPER::PHE2* transgene increased the level of *PHE2* expression in the seed.

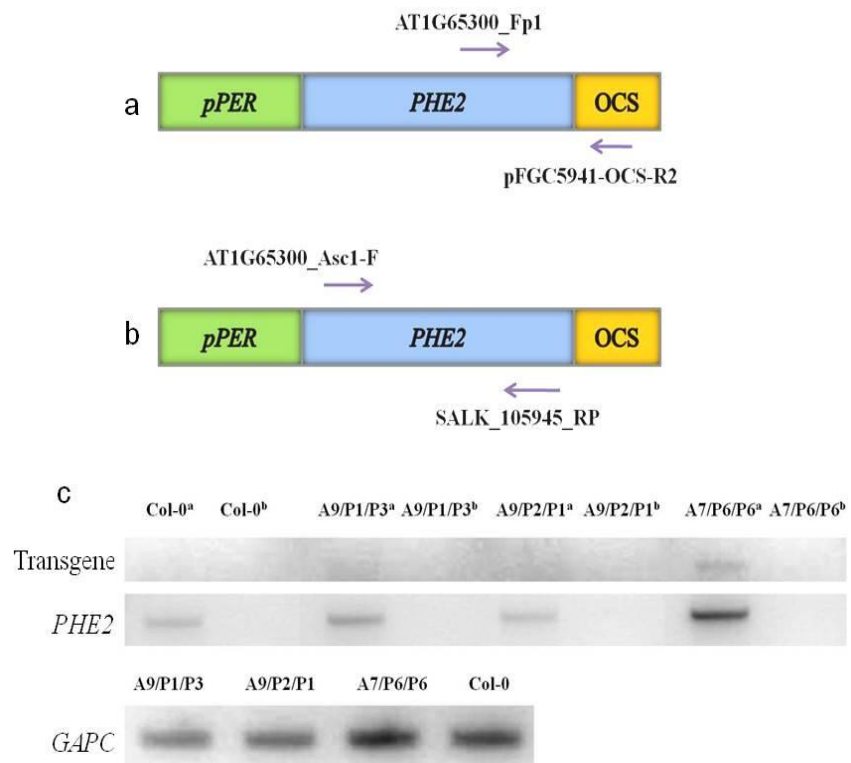


Figure 3.10 Semi-quantitative RT-PCR analysing *pPER::PHE2* transcript levels of three homozygous, single insertion lines. (a) Primer orientation for *PHE2* transgene amplification. (b) Primer orientation for total transcript amplification. (c) Semi-quantitative RT-PCR on RNA extracted from 5 DAP siliques of the over-expression lines and Col-0. ^acDNA, ^b RNA as a negative control. The *GAPC* gene was used as a control.

3.2.3.2 Analysis of *pPER::CYCD4;1* plants

3.2.3.2.1 Seed size analysis

To determine whether over-expression of *CYCD4;1* in the endosperm altered seed size, four independent *pPER::CYCD4;1* transformation lines that were homozygous for a single insertion and showed promising high seed weights within the primary transformants were analysed for sustained seed weight increase. All lines produced seed that was either significantly heavier or lighter than wild type (Figure 3.11a). Two *pPER::CYCD4;1* transgenic lines produced significantly lighter seeds than wild type; seed harvested from A1/P9/P4 and A2/P4/P3 were respectively 10% ($p < 0.001$) and 5% ($p = 0.022$) lighter than wild type. These plants produced an average of 50 seeds/silique indicated that fertility of these was not significantly different to wild type ($p = 0.728$ and 0.775 , respectively). In contrast, both *pPER::CYCD4;1*-A1/P10/P8 and *pPER::CYCD4;1*-A1/P4/P9 maintained an increased seed weight an increase: , 26% ($p < 0.001$) and 20% ($p < 0.001$) respectively, compared to wild type. However, fertility was substantially reduced in both lines, with the number of seeds per silique respectively, 46% ($p < 0.001$) and 57% ($p < 0.001$) compared to wild type (Figure 3.11b). Therefore, the observed increase in seed weight over wild type was most likely an indirect result of reduced seed number.

To test whether the changes in seed size observed in *pPER::CYCD4;1* lines were correlated with an increase in endosperm proliferation, two *pPER::CYCD4;1* lines were selected for embryo sac area measurement, *pPER::CYCD4;1*-A1/P4/P9 and *pPER::CYCD4;1*-A1/P9/P4 (Figure 3.12). Although *pPER::CYCD4;1*-A1/P4/P9 plants produced heavier seed than wild type, the embryo sac area at 6 DAP showed no significant difference to Col-0 at the same age ($p = 0.481$). In addition, embryo sac area of *pPER::CYCD4;1*-A1/P9/P4 line that produced lighter seeds, showed no significant difference to wild type at 6 DAP ($p = 0.738$). Similar to *pPER::PHE2* plants phenotype that the final seed weight includes contributions from the embryo, endosperm and the seed coat (integument layer) and not only the endosperm.

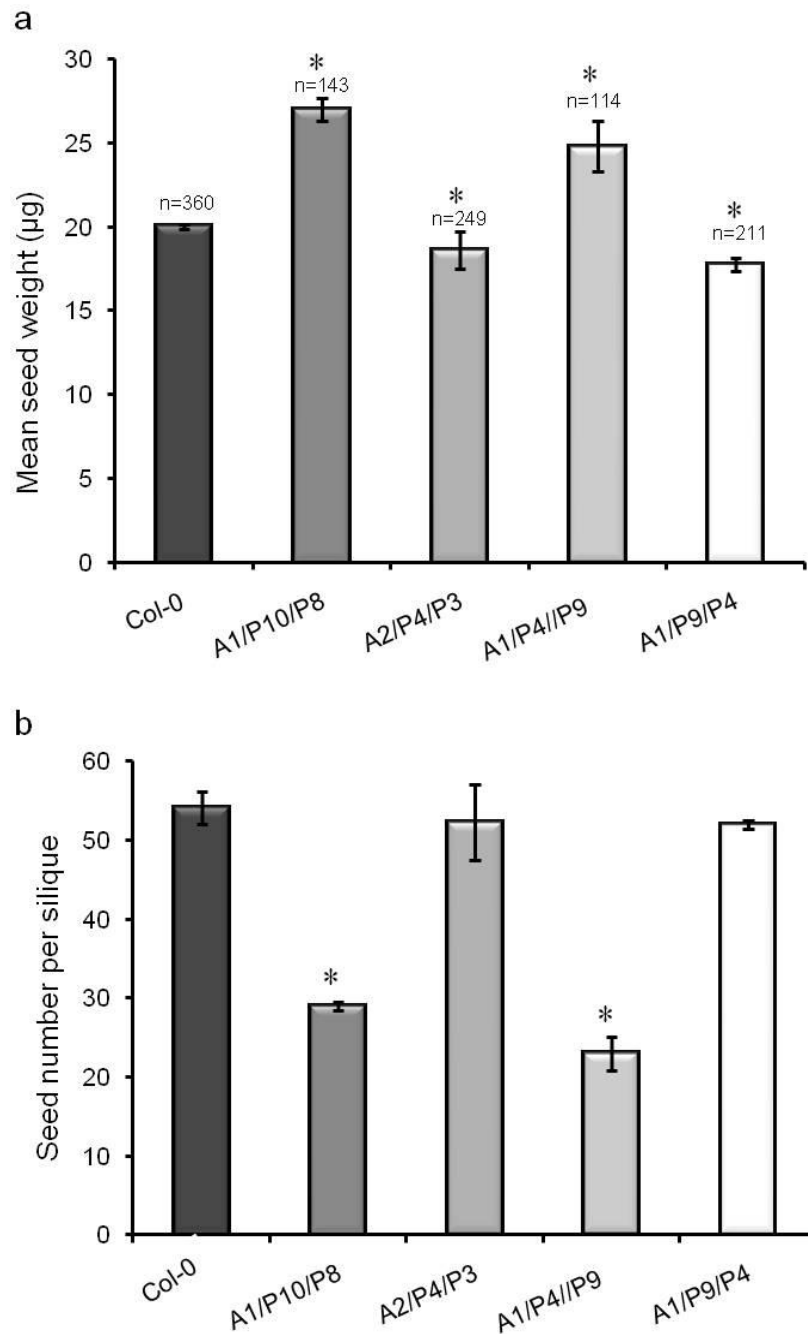


Figure 3.11 Seed weight analysis of four independent, homozygous, single insertion *pPER::CYCD4;1* plants compared to Col-0. (a) Mean seed weight; n = number of seeds examined. (b) Seed set per silique (n = 5 siliques/ 3 plants). For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *. Error bar = S.E.M.

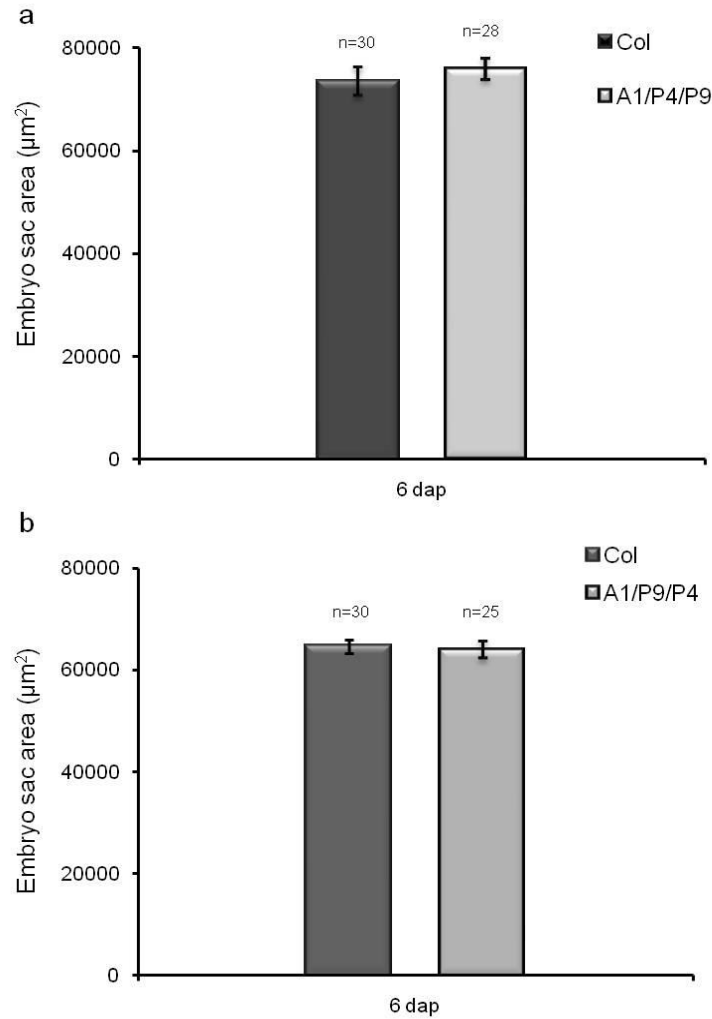


Figure 3.12 Embryo sac area measurement of wild type and two independent, homozygous, single insertion *pPER::CYCD4;1* plants at 6 DAP. (a) Mean embryo sac area of *pPER::CYCD4;1*-A1/P4/P9 and Col-0. (b) Mean embryo sac area of *pPER::CYCD4;1*-A1/P9/P4 and Col-0. For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *; n = number of seeds examined. Error bar = S.E.M.

3.2.3.2.2 Expression of *CYCD4;1* in *pPER::CYCD4;1* plants

Semi-quantitative RT-PCR was carried out on cDNA prepared from siliques to determine whether the *pPER::CYCD4;1* transgene was both expressed and capable of enhancing the level of *CYCD4;1* mRNA expression in the endosperm. The procedure was essentially as described above for *pPER::PHE2* (Figure 3.13a; Figure 3.13b). Although, the *pPER::CYCD4;1*-A1/P4/P9 line tested positive for the presence of the *pPER::CYCD4;1* transgene, the A1/P4/P9 transgenic line showed no expression of *CYCD4;1* in the semi-quantitative RT-PCR, indicating no increase in level of the *CYCD4;1* mRNA (Figure 3.13c).

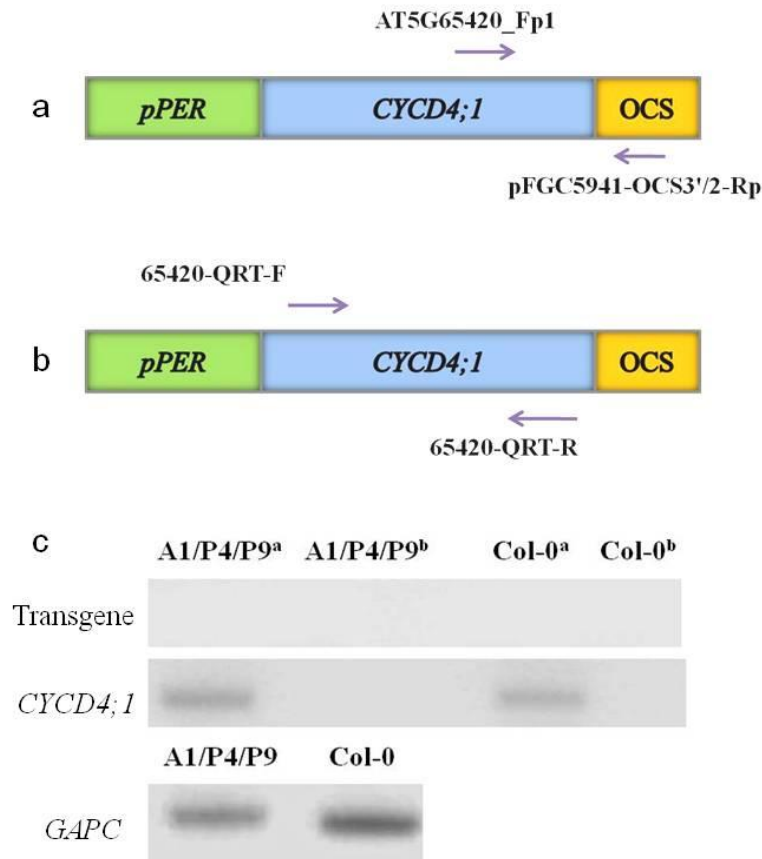


Figure 3.13 Semi-quantitative RT-PCR analysing *pPER::CYCD4;1*-A1/P4/P9 transcript level. (a) Primer orientation for *CYCD4;1* transgene amplification. (b) Primer orientation of total transcripts amplification. (c) Semi-quantitative RT-PCR on RNA extracted from 5 DAP siliques of the *pPER::CYCD4;1*-A1/P4/P9 line and Col-0. ^a cDNA, ^b RNA as a negative control. The *GAPC* gene was used as a control.

3.2.3.3 Analysis of *pPER::GAI* plants

3.2.3.3.1 Seed size analysis

To determine whether over-expression of *GAI* in the endosperm altered seed size, two independent *pPER::GAI* transformation lines that were homozygous for a single insertion and showed promising high seed weights in the previous generation were analysed for sustained seed weight increase. One *pPER::GAI* line produced seed that was significantly heavier than wild type and the other line was similar to wild type (Figure 3.14a). The A10/P4/P8 plants produced significantly heavier seeds than wild type (13%; $p < 0.001$). This plant produced an average of 50 seeds/silique indicated that fertility of these was not significantly different to wild type ($p = 0.984$). In contrast, A11/P5/P8 plants had a similar

seed weight to wild type ($p = 0.109$). However, fertility was higher in *pPER::GA1*-A11/P5/P8 plants with the number of seed per silique 22% ($p = 0.023$) compared to wild type (Figure 3.14b). A trade-off exists between seed number and weight; hence it was possible that the increased seed number in A11/P5/P8 masked any increase in seed weight due to GA1 over-expression.

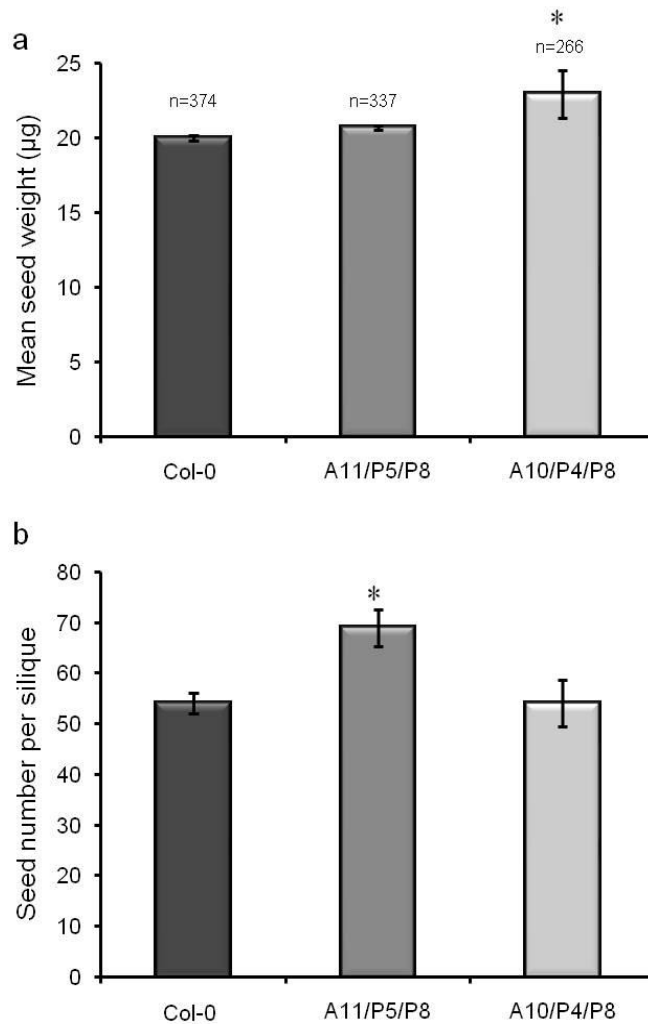


Figure 3.14 Seed weight analysis of two independent, homozygous, single insertion *pPER::GA1* plants. (a) Mean seed weight; n = number of seeds examined. (b) Seed set per silique (n = 5 siliques/ 3 plants). For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *. Error bar = S.E.M.

To test whether the changes in seed size observed in *pPER::GA1* lines were correlated with an increase in endosperm proliferation, the two *pPER::GA1* lines were subjected for embryo sac area measurement, *pPER::GA1*-A10/P4/P8 and *pPER::GA1*-A11/P5/P8 (Figure 3.15). The embryo sac area of *pPER::GA1*-A10/P4/P8 seeds was significantly

larger (15%; $p < 0.001$) than wild type, that translated into a heavier seeds compared to wild type. In contrast, although *pPER::GAI*-A11/P5/P8 plants produced similar seed to wild type, the embryo sac area at 6 DAP showed a significant increase to Col-0 at the same age (7%; $p = 0.010$).

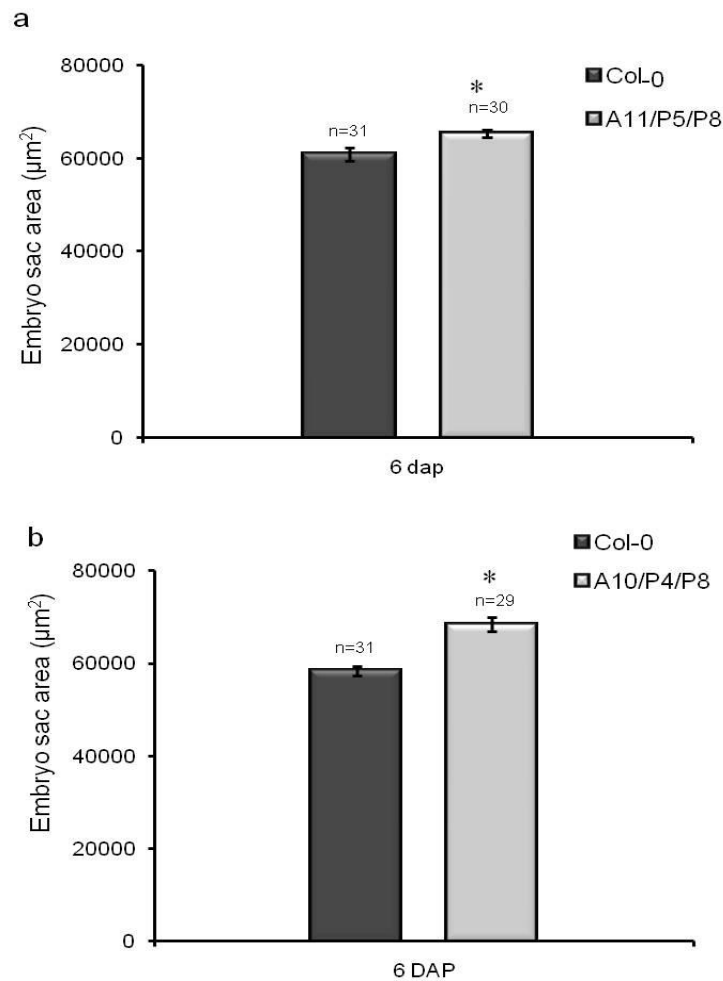


Figure 3.15 Embryo sac area measurement of wild type and two independent, homozygous, single insertion *pPER::GAI* plants at 6 DAP. (a) Mean embryo sac area of *pPER::GAI*-A11/P5/P8 and Col-0. (b) Mean embryo sac area of *pPER::GAI*-A10/P4/P8 and Col-0. For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *; n = number of seeds examined. Error bar = S.E.M.

3.2.3.3.2 Expression of *GAI* in *pPER::GAI* plants

Semi-quantitative RT-PCR was carried out on cDNA prepared from siliques to determine whether the *pPER::GAI* transgene was both expressed and capable of enhancing the level of *GAI* mRNA expression in the endosperm. The procedure was essentially as described

above for *pPER::PHE2* (Figure 3.16a; Figure 3.16b). Two independent transformation lines were analysed. Whilst all tested positive for the presence of *pPER::GAI* transgene there was a spectrum of *GAI* expression levels in the semi-quantitative RT-PCR (Figure 3.16c). Only the A11/P5/P8 line produced detectable levels of *GAI* mRNA, providing strong supporting evidence that the *pPER::GAI* transgene increased the level of *GAI* expression in the seed.

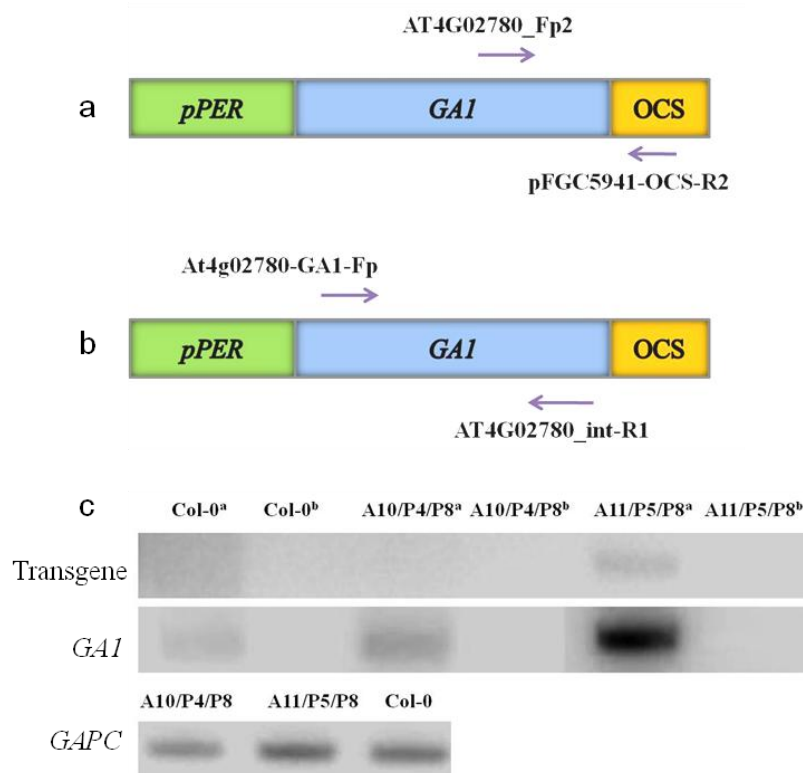


Figure 3.16 Semi-quantitative RT-PCR analysing *pPER::GAI* transcript levels. (a) Primer orientation for *GAI* transgene amplification. (b) Primer orientation of total transcripts amplification. (c) Semi-quantitative RT-PCR on RNA extracted from 5 DAP siliques of the *pPER::GAI* transgenic lines and *Col-0*. ^a cDNA, ^b RNA as a negative control. The *GAPC* gene was used as a control

3.2.3.4 Seed size analysis of *pPER::GA20-Oxidase 5* plants

To determine whether over-expression of *GA20OX5* in the endosperm altered seed size, three independent *pPER::GA20OX5* transformation lines that were homozygous for a single insertion and showed promising high seed weight in previous generation were analysed for increased in the seed weight. All lines tested were no different to the wild type as far as seed set was concerned, they produced 50 seeds/silique (ANOVA followed by Tukey's multiple comparisons test, $p = 0.164$) (Figure 3.17b). Although, there was no

significant difference in seed fertility compared to the wild type, *pPER::GA20OX5-1/2/A/5* plants produced lighter seeds than Col-0 ($p = 0.022$). In contrast, the other two transgenic plants, *pPER::GA20OX5-1/11/A/6* and *pPER::GA20OX5-2/2/C/6*, were similar to wild type in seed weight ($p = 0.241$ and 0.571 , respectively) (Figure 3.17a).

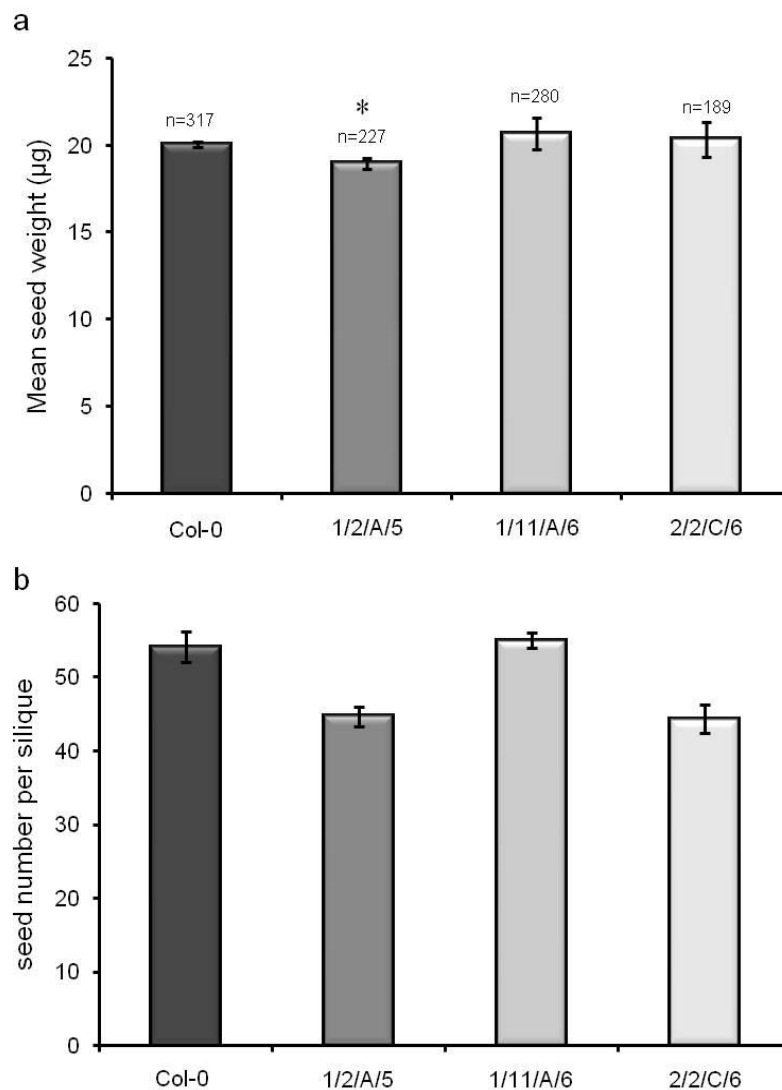


Figure 3.17 Seed weight analysis of three independent, homozygous, single insertion *pPER::GA20OX5* plants compared to Col-0. (a) Mean seed weight; n = number of seeds examined. (b) Seed set per silique (n = 5 siliques/ 3 plants). For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *. Error bar = S.E.M.

3.3 Discussion

There are several approaches to engineer seed size; we have demonstrated that at least in *A. thaliana* seed size is regulated by parent of origin effects in the interploidy crosses. There are two components to manipulating seed size, one is through integument led growth as demonstrated in the *megaintegumenta* (*mnt*) mutants (Schruff *et al.*, 2006) and the other is by increasing number of endosperm nuclei or altering the timing of cellularisation of the endosperm (Scott *et al.*, 1998). Even though there has been a surge of data generated from endosperm directly via endosperm tissue profiling or indirectly through the profiling expression of the interploidy crosses (Day *et al.*, 2008; Goldberg and Harada dataset (<http://seedgenenetwork.net/arabidopsis>); Tiwari *et al.*, 2010), very little is known about the function of these genes and their role in endosperm development. Our over-expression study with the candidate genes aimed to determine the role of these genes in endosperm development and the feasibility of increasing seed size via single gene manipulation. By transferring our knowledge of these genes to crop plants we can build a strategy to employ these genes to engineer larger seeds and higher yield in other crop plants. The objective of the study in this chapter was to assess the role of such candidate genes that are involved in endosperm development following knock in strategy by using either 35S promoter or endosperm specific promoter (*pPER*). The phenotype of transgenics was assessed for increased seed size and some positive outcomes were observed.

3.3.1 Expression of candidate genes

All of the candidate genes shortlisted from the microarray were found to be expressed either exclusively or generally in the endosperm as outlined in Section 3.2.1. *PHE2*, *AGL40*, *AGL62* and *AGL45* are highly expressed in paternalized endosperm. These genes belong to the same family of transcription factors (MADS-box protein) which are key regulators of several plant development processes (Alvarez-Buylla *et al.*, 2000a; Alvarez-Buylla *et al.*, 2000b; Parenicova *et al.*, 2003). Other interesting genes, which belong to various families like, *CKX2*, *MAPK10*, *E2L2*, *CYCD4;1*, *CYCD4;2*, *GA1*, *GA20-oxidase 5*, *GA-regulated* and *AT5G46950* are also highly expressed in paternalized endosperm and thus can be hypothesised to be positively associated with endosperm proliferation.

3.3.2 Over-expression of the candidate genes from the 35S promoter

The experimental system for the KI strategy (Section 3.2.2) predicted that over-expression of a growth promoter gene from the CaMV 35S constitutive promoter, which is highly

expressed in the vegetative parts of the plant gave rise to thicker stems, and both larger leaves and flowers (*PHE2*, *AGL40*, *AT5G46950*, *E2L2*, *MAPK10* and *CKX2*).

3.3.2.1 Over-expression of *E2L2*, *MAPK10* and *CKX2* under the 35S promoter results in vegetative abnormalities

KI lines of *PHE2*, *AGL40* and *AT5G46950* yielded plants that were indistinguishable from wild type plants (Table 3.4). In contrast, plants with obvious abnormal phenotypes were found among the KI lines of the *E2L2*, *MAPK10* and *CKX2* genes.

E2L2 protein is up regulated in large seeds according to our microarray data (Tiwari *et al.*, 2010; Table 3.2); we therefore expected large vegetative and reproductive organs when the transcript was over-expressed from the 35S promoter. Over-expression of this gene by 35S promoter however, resulted in plants with abnormal woody and curly stems, and flowers that failed to produce selfed seed (Figure 3.4). These flowers were not completely sterile since pollination with wild type pollen resulted in a small number of plump seeds (Figure 3.7b). *E2L2* belongs to E2F family of transcription factor that control cell cycle transitions in multicellular organisms. Ramirez-Parra *et al.*, (2004) investigated over-expression of *A. thaliana* AtE2Ff and found reductions in the size of differentiated cells in roots and hypocotyls organs. In addition, the authors suggested that AtE2Ff had a repression effect on cell wall biosynthesis genes during cell elongation in differentiated cells, which means that it is a molecular regulator of other genes. These phenotypes do not appear consistent with *E2L2* encoding a growth promoter as predicted from its expression profile in our microarray data.

Similar to *E2L2*, the *MAPK10* gene showed high expression in paternalized seeds (Tiwari *et al.*, 2010; Table 3.2). *MAPK10* encodes a mitogen-activated protein (MAPK10). Over-expression of *MAPK10* using the 35S promoter resulted in longer stems than the wild type, long pistil and very short silique exhibiting low fertility (Figure 3.5 and 3.7d). These phenotypes are arguably consistent with *MAPK10* encoding a growth promoter as expected from the expression profile of the gene in our microarray data.

The transcript of *CKX2* gene was also more abundant in large seeds (Table 3.2). *CKX2* encodes a cytokinin oxidase/dehydrogenase protein (CKX2) which regulates cell division and proliferation by affecting cytokinin hormone activity (Werner *et al.*, 2003; 2006). The 35S::*CKX2* transgenic plants were stunted, 5 days earlier in flowering compared to the wild type and produced inflorescences composed of small flowers with long pistil (Figure

3.6). On its own, the plant rarely produced seeds, but assisted pollination resulted in seeds, which were abnormal, and mostly abortive (Figure 3.7e). Evidence reported by Wernr *et al.* (2003) showed that over-expression of *AtCKX2* resulted in pleiotropic developmental changes during all phases of the growth cycle; the transgenic plants were stunted, however, and produced a reduced number of otherwise normal flower. These phenotypes, and those found in the present work, do not appear consistent with *CKX2* encoding a growth promoter as predicted from its expression profile in our microarray data.

As reviewed in the literature, the CaMV 35S promoter is considered a strong constitutive promoter (Guilley *et al.*, 1982; Odell *et al.*, 1985; Jefferson *et al.*, 1987; Cornejo *et al.*, 1993; Steinbrecher 2002). The promoter is widely used to express the transgenes in plants. Although 35S is a strong promoter with a very wide expression pattern, constitutive expression of the foreign genes may cause abnormalities to the host plants, such as sterility, retarded development, abnormal morphology, yield penalty, altered grain composition or transgene silencing (Cai *et al.*, 2007). This last author, and many others, have expressed the view that limiting transgene expression to target tissues (tissue specific promoters) or limiting the time when they are active (inducible promoters) may circumvent the problem of unnecessary costs to the plant that may reduce yields.

Zhang (2003) suggested that over-expression phenotypes be classified into two types: hypermorphs and neomorphs. Hypermorphs are over-expressers that cause an increase in otherwise normal gene function, while neomorphs are over-expressors or gain-of-function mutations that cause novel gene function. Therefore, one of limitations of the over-expression technique is the possibility of creating neomorphic alleles that result in phenotypes that are apparently inconsistent with the proposed function of the target gene (Zhang, 2003). This may explain the phenotypes produced by some of the over-expression lines in the present work.

This present study focused on endosperm development and its correlation with final seed size. However because the phenotypes observed from the 35S transgenic plants were difficult to interpret despite showing some degree of over-proliferation, it was decided to create transgenic lines with expression of the candidate genes driven by an endosperm specific promoter. This idea was supported by Zhang (2003) in his review about over-expression analysis of plant transcription factors. Over-expression of transcription factors can result in phenotype that are difficult to interpret (Helariutta *et al.*, 2000), because they are key regulators of many developmental and biochemical processes in plants. Therefore,

in order to obtain instructive phenotypes, it is often necessary to express the transcription factors in very specific tissues and/or during specific developmental stages.

3.3.3 Over-expression of the candidate genes by using *pPER* promoter

Over-expression using 35S promoter was unsuccessful to achieve our aim because the transgenic plants showed some degree of over-proliferation in all plant parts and inconsistent phenotype; therefore we decided to use the *pPER* promoter to over-express the target genes, this circumvents the problem of nonspecific expression in other tissues and limits the expression to a great extent to the developing endosperm (Section 3.2.3).

3.3.3.1 Over-expression of genes which positively associated with endosperm growth has the potential to increase seed size

The *pPER::PHE2*, *pPER::AGL40*, *pPER::AGL45*, *pPER::GA20OX5*, *pPER::GA1*, *pPER::GA-regulated*, *pPER::CYCD4;1* and *pPER::CYCD4;2* constructs were introduced into *A. thaliana* with the aim of increasing endosperm growth and development and increasing seed size. The endosperm-specific promoter *pPER* was selected for the purpose (Tiwari *et al.*, 2006). The first screening of for overall phenotypes including vegetative and reproductive parts was done in a previous study and only the constructs that exhibited a promising phenotype in transgenic plants were analysed further in the present work.

Three *pPER::PHE2* lines (A9/P2/P1, A9/P1/P3 and A7/P6/P6), had heavier seed than wild type, but also displayed varying levels of seed set (Table 3.5; Figure 3.8). Since a trade-off exists between seed number and seed size, where a reduction in seed number results in an increase in seed weight, and *vice versa* (Venable, 1992), it was important to measure fertility in the lines before drawing any conclusions about the efficacy of the transgene. Analysis of A7/P6/P6 for seed number per silique, a good measure of fertility, revealed normal levels of seed fertility, suggesting that the increase in seed size observed in this line was not due to reduced seed set. In contrast, the heavy seed weight of the other two lines *pPER::PHE2*-A9/P2/P1 and *pPER::PHE2*-A9/P1/P3 was associated with a reduction in seed number. It was therefore not safe to assume that the increased seed weight in these lines was due to the transgene.

Logically, an increase in seed weight due to the transgene increasing endosperm size should be detectable during early seed development. Embryo sac area measurement at 6 DAP for *pPER::PHE2*-A9/P2/P1 and *pPER::PHE2*-A7/P6/P6 revealed that *pPER::PHE2*-

A9/P2/P1 seeds were indeed 13% larger than the control. However, *pPER::PHE2*-A7/P6/P6 seeds were a similar size to wild type. Thus there was no correlation between seed size increase and embryo sac area. One explanation for the discrepancy is that the embryo sac area measures endosperm size only, but final seed weight includes contributions from the embryo, endosperm and the seed coat. In addition, we found that only the A7/P6/P6 line produced detectable levels of *PHE2* mRNA confirmed by RT-PCR (Figure 3.10c), providing strong supporting evidence that the *pPER::PHE2* transgene increased the level of *PHE2* expression in the seed. This finding suggests that introduction of the *pPER::PHE2* transgene into plants has altered seed development and as a result increased final seed size.

Like *PHE2*, the transcript of *CYCD4;1* gene was also up-regulated in large seeds but its expression was also reported in all plant tissues (Table 3.2). Only two *pPER::CYCD4;1* lines namely: *pPER::CYCD4;1*-A1/P10/P8 and *pPER::CYCD4;1*-A1/P4/P9, produced seed that was significantly heavier than wild type with an increase of 26% and 20% respectively (Figure 3.11a). However, both lines showed low seed fertility (Figure 3.11b) suggesting that the increased seed weight in these lines was not due to the transgene. Further analysis of *pPER::CYCD4;1*-A1/P4/P9 plants showed that the embryo sac area of the transgenic plants was similar to wild type at 6 DAP and the expression of *CYCD4;1* mRNA was not detected by RT-PCR (Figure 3.13c). Therefore, we can suggest that the observed increase in seed weight over wild type was most likely an indirect result of reduced seed number.

GAI transcript was 5-fold more abundant in paternalized 2x X 6x seed (Table 3.2). Variation was found among different transgenic lines containing *pPER::GAI* construct, but importantly *pPER::GAI*-A10/P4/P8 and *pPER::GAI*-A11/P5/P8 showed significant increase either in seed size or seed set compared to wild type. The A10/P4/P8 plants produced a 13% heavier seeds with similar fertility levels to wild type and had 15% larger embryo sac areas (Figures 3.14 and 3.15). In contrast A11/P5/P8 plants showed similar seed size with a 22% higher seed set and a 7% larger embryo sac compared to wild type. The RT-PCR revealed that only A11/P5/P8 line showed detectable levels of *GAI* mRNA, providing strong supporting evidence that the *pPER::GAI* transgene increased the level of *GAI* expression in the seed. Therefore, this increase in seed weight over wild type was not associated with a reduction in seed number but is due to altered endosperm.

GA20-oxidase5 transcript was also up regulated in large seeds and its expression was present in endosperm and root endodermis, as reported by different microarrays (Table 3.2). However, plants transformed with the *pPER::GA20OX5* construct showed no consistent alterations in seed size or seed fertility. The failure to increase seed size by transforming plants with *pPER::GA20OX5* may be explained in a number of ways. However, since we did not determine whether the construct increased the level of *GA20OX5* at the critical developmental stage (4-5DAP) lack of appropriate gene expression remains the most likely cause of failure.

In conclusion, over-expression using *pPER* promoter has the potential to increase seed size as illustrated by the positive results obtained for the *pPER::PHE2* and *pPER::GAI* constructs. However, even though the promoter used was both endosperm specific and active during the relevant development phase of the seed, the levels of over-expression were relatively modest, suggesting that greater potential exists to increase seed size by employing a stronger promoter.

3.3.4 Future work

The *pPER::AGL40* and *pPER::AGL45*, *pPER::PHE1* and *pPER::AGL28* (Hughes, 2009) transgenes did not result in any significant change in seed size, whilst the present study showed that over-expression of *PHE2* under the same promoter potentially yielded heavier seeds. Since the proteins of these genes appear to interact with each other (de Folter *et al.*, 2005; Section 3.2.1.1), these may be required to operate within multi-protein complex to alter endosperm development. Over-expression of a single gene in this case would therefore be insufficient to alter endosperm size. This idea is supported by our findings from knockout mutant analysis (Chapter 6) that these genes function as complexes to affect endosperm development. Future work should focus on combining transgenes into the same plant and testing the impact on seed size.

Chapter 4: Phenotypic characterization of *phe1*

4.1 Introduction

In the previous chapters we discussed the correlation between seed growth and endosperm growth in crop plants and the potential for increasing seed size by manipulating endosperm proliferation. *PHERES1* (*PHE1*) is one of the six known imprinted genes in *Arabidopsis thaliana* (*A. thaliana*) and the only gene reported to be highly expressed from the paternal genome and repressed from the maternal genome (Kohler *et al.*, 2005). According to the parental conflict model for the evolution of genomic imprinting, this pattern of expression would indicate that *PHE1* encodes a protein capable of promoting endosperm growth and development (Haig and Westoby, 1991). Although, *PHE1* had been subjected to extensive molecular genetic analysis (Kohler *et al.*, 2003b; Kohler *et al.*, 2005; Makarevich *et al.*, 2008; Villar *et al.*, 2009), there were no published investigations addressing whether *PHE1* encodes an endosperm growth promoter. Since *PHE1* was one of the candidate genes within the list generated by our microarray experiments (Section 3.2.1; Tiwari *et al.*, 2010), experiments were undertaken to address the role of *PHE1* in the endosperm.

4.1.1 *PHE1* is a MADS box transcription factor

PHE1 (AGL37) encodes a MADS-domain protein of the type I class (Alvarez-Buylla *et al.*, 2000a). The *PHE1* protein contains 279 amino acids and has a predicted molecular weight of 32 kD (Kohler *et al.*, 2003b). The *A. thaliana* genome contains 107 potential MADS box genes (Parenicova *et al.*, 2003). Although the function of many of these genes is unknown, some have been allocated roles in different aspects of plant development, such as plant reproduction, flowering time, floral organogenesis, floral meristem identity, fruit formation, seed pigmentation and endothelium development (Ng and Yanofsky, 2001; Nesi *et al.*, 2002; Becker and Theissen, 2003). In addition, de Folter *et al.* (2005) reported that *PHE1* protein was within a list of co-regulated and interacting MADS box transcription factors that belong to the type I MADS group and showed an interaction with AGL62, AGL40 and AGL28 proteins in a yeast-two hybrid study (Chapter 6; Figure 6.1). Day *et al.* (2008) also showed that *PHE1* was highly expressed in the proliferative phase of endosperm development and that its protein had a strong interaction with AGL62 and AGL40 separately in tissue specific analysis. Kohler *et al.* (2003b) identified a close homolog of *PHE1* (72% amino acid sequence identity) named *PHE2* located ~10 kb from *PHE1* on Chromosome 1. The authors analyzed *PHE2* expression during seed

development and showed that it is expressed in a similar pattern as *PHE1* in wild type plants, but at an eightfold lower level compared with *PHE1*.

PHE1 is the only imprinted gene identified in *A. thaliana* that is expressed paternally and suppressed maternally (Kohler *et al.*, 2005). In general, imprinting involves either nucleic acid-based mechanisms, such as small RNA molecules (Bartel, 2004; Zhang *et al.*, 2007) or DNA methylation (Reik *et al.*, 2001; Saurin *et al.*, 2001; Henderson and Jacobsen, 2007), as well as post-translational modifications of histone proteins (Jenuwein and Allis, 2001).

DNA methylation in plants affects cytosine residues adjacent to guanines (CG), but may also occur on cytosines at CNG (where N is any nucleotide) sites and in the asymmetrical sequence CHH (where H is any nucleotide excluding G) (Vaillant and Paszkowski, 2007). Imprinting control of some genes in plants requires active DNA demethylation catalyzed by DNA glycosylases (Reik *et al.*, 2001; Saurin *et al.*, 2001; Henderson and Jacobsen, 2007) such as DEMETER (DME) which acts antagonistically to the maintenance methyltransferase 1 (MET1) that methylates *MEDEA* (*MEA*) in the promoter and 3' regions of the gene (Xiao *et al.*, 2003).

In histone modifications, the histone-tails are modified by the addition of various covalent attachments, including methyl, acetyl and phosphate groups, (Jenuwein and Allis, 2001). An example is the methylation of lysine 27 on histone 3 (H3K27me3) by histone methyltransferases (HMTases) (Lachner and Jenuwein, 2002; Lachner *et al.*, 2003) mediated by a Polycomb group protein (PcG) complex (Czermin *et al.*, 2002; Muller *et al.*, 2002). The PcG proteins are regulators of genomic programs that assemble as high molecular weight complexes and have histone methyltransferase activity targeting lysine 27 of histone H3 (Schwartz and Pirrotta, 2007). The action of PcG proteins was firstly considered to regulate only homeotic genes. More recently, the genome-wide identification of PcG target genes reveals that PcG proteins are global transcriptional repressors that shut down alternative genetic programmes not required in a particular cell type (Kohler *et al.*, 2010).

Both DNA methylation and histone modification events occur in the central cell and sperm cell and are responsible for determining the structure of chromatin and its transcriptional competence (Huh *et al.*, 2007; Vaillant and Paszkowski, 2007; Huh *et al.*, 2008). In plants epigenetic marks that differentiate paternal and maternal imprinted alleles need not be reset

for each generation, because the endosperm, formed by a separate fertilization event, does not transmit its genome to the next generation (Huh *et al.*, 2008).

The repression of *PHE1* maternal allele expression is caused by the *FERTILIZATION INDEPENDENT SEED* (*FIS*) Polycomb group complex, which contains *MEDEA* (*MEA*), *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*), *FIS2* and *MULTI-COPY SUPPRESSOR OF IRA1* (*MSI1*) (Figure 4.1) (Kohler *et al.*, 2005; Makarevich *et al.*, 2006). The FIS-PcG complex, including *MEA* itself, repressed the paternal *MEA* allele via trimethylation of H3K27 (Baroux *et al.*, 2006; Gehring *et al.*, 2006). Also, the FIS-PcG complex is active in the female gametophyte as well as in the endosperm that avoided precocious *PHE1* expression before fertilization and restricts *PHE1* expression in the chalazal endosperm after fertilization (Kohler *et al.*, 2005). Both Kohler *et al.* (2003b) and Makarevich *et al.* (2006) reported that the FIS complex is directly associated with the *PHE1* locus and that FIS suppressive activity is correlated with H3K27me3 modification at *PHE1*. In addition, the PcG protein *MEA* controls the expression of *PHE1* because *PHE1* is paternally expressed and maternally repressed and this maternal repression of *PHE1* breaks down in maternal *mea* mutant seeds (Kohler *et al.*, 2005). Also, the up-regulation of *PHE1* in *mea* mutants is in part responsible for the *mea* mutant phenotype that can be eased by reducing *PHE1* expression (Kohler *et al.*, 2003b; Makarevich *et al.*, 2006).

Makarevich *et al.* (2008) reported that the DNA methylation status of a repetitive sequence in a regulatory region located in the 3' region about 2.5 kb downstream from the *PHE1* locus affected the imprinted expression of the gene. The authors also showed that the 3' repetitive sequence of *PHE1* is usually methylated in leaf tissue of wild type plants. However, in *met1* mutants, DNA methylation of the 3' repetitive sequence was lost causing silencing of the paternal *PHE1* allele, while, a transgene lacking the 3' region did not show imprinted expression of *PHE1* (Makarevich *et al.*, 2008). Therefore, DNA methylation of the repeat is necessary for active expression of the paternal *PHE1* allele. In summary, the imprinting of *PHE1* evolved both FIS PcG complex repression activities (Kohler *et al.*, 2005) as well as on a distantly located demethylated repeat region (Makarevich *et al.*, 2008).

Recently, Luo and co-workers (2008) reported that UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26) participated in repressing the maternal allele of *PHE1* expression and may act in similar pathway to the FIS complex. The authors showed that *PHE1* expression is high

in *Atubp26* mutant siliques compared to wild type and is associated with the loss of H3K27me3 as in a *mea* mutant (Makarevich *et al.*, 2006).

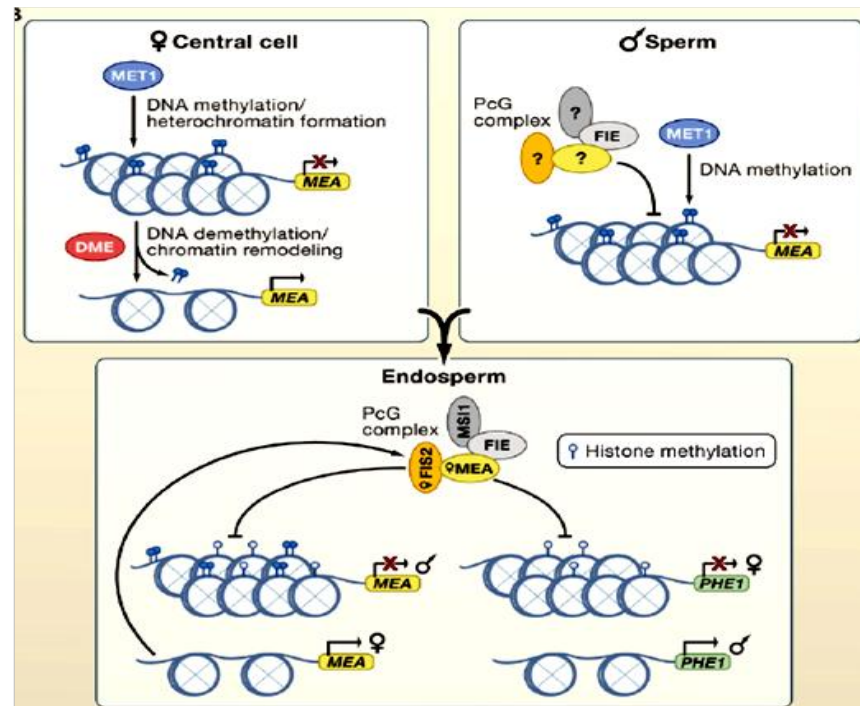


Figure 4.1 Model for the imprinting of *MEA* and *PHE1* (adapted from Huh *et al.* 2008). The maternal *MEA* allele is demethylated and activated by DME DNA glycosylase in the central cell. Silencing of the paternal *MEA* allele in sperm may be due to methylation by cytosine MET1, as well as histone H3K27 methylation by a PcG complex that includes the FIE Polycomb group protein. Upon fertilization, maternally expressed *MEA* participates in a PcG complex that represses targets such as the paternal *MEA* allele and the maternal *PHE1* allele via histone H3K27 methylation. Maternal *MEA* that is not repressed is continuously expressed in the early endosperm, replenishing the PcG complex. Consequently, *MEA* is maternally expressed while *PHE1* is paternally expressed in the endosperm. Only one of the two maternal alleles is shown in the central cell and endosperm.

4.2 *PHE1* expression is associated with endosperm over-proliferation

As mentioned earlier (Chapter 1), in *A. thaliana*, interploidy crosses between a diploid seed parent and a tetraploid pollen parent (2x X 4x) result in enlarged seed. This paternal excess phenotype is associated with increased and prolonged proliferation of the peripheral endosperm, overgrowth of chalazal endosperm and surrounding nodules, and a delay in endosperm cellularisation (Scott *et al.*, 1998). The reciprocal cross (4x X 2x) causes the

opposite effect on seed size and endosperm development. The gene expression underlying paternal and maternal excess phenotypes was subjected to analysis by transcript profiling (Section 1.6; Tiwari *et al.*, 2010). This identified genes that were up-regulated in a paternalized seed and therefore associated with endosperm over-proliferation, which was used to derive a shortlist of genes (Chapter 3; Table 3.1) for functional verification in the present study. *PHE1* was one of the shortlisted genes.

The *PHE1* transcript was more abundant in paternalized seeds than in ‘balanced’ seed produced by a 2x X 2x cross (Tiwari *et al.*, 2010). *PHE1* expression levels were fivefold higher in both paternal excess crosses (2x X 4x and 2x X 6x) compared to 2x X 2x cross (Chapter 3; Table 3.2). In addition, Table 3.2 shows that *PHE1* expression was restricted to endosperm as determined by different microarray experiments.

According to Tiwari *et al.* (2010), *PHE1* and *PHE2* share the same Affymetrix probeset and therefore either transcript may have been responsible for the signal. Also by analysing the expression profile of *PHE1* and *PHE2* by applying a quantitative real-time PCR (qRT-PCR) using gene-specific primers showed similarity in expression in both genes, which indicate that both *PHE1* and *PHE2* are likely to be up-regulated in seeds with paternal excess or a paternalizing *fisIX2x* mutation (Tiwari *et al.*, 2010).

As *PHE1* is paternally expressed, the parental conflict theory predicts that *PHE1* encodes a growth promoter that should act to increase endosperm proliferation. As explained earlier, the KO strategy was designed to test whether inactivation of a target gene produced the expected phenotypic change in endosperm growth and development. For a proposed growth promoter such as *PHE1*, inactivation would hypothetically result in reduced endosperm proliferation and an associated reduction in seed mass. This chapter describes the results of experiments with a KO allele of *PHE1*.

4.2 Results

4.2.1 Verifying the KO status of the *phe1.1* mutant alleles

In order to verify the genotype of the insertion line before engaging in further experiments, plants were grown from seeds donated by Dr. Claudia Kohler (ETH Zurich) as a homozygous line which originated from the Cold Spring Harbor Laboratory (CSHL). As reported in previous studies the *A. thaliana* mutant *phe1.1* (CSHL line ET189) contains a Dissociation (Ds) transposon insertion in *PHE1*, and this insertion results in complete loss

of function but there were no recorded phenotypes (Kohler *et al.*, 2005; Josefsson *et al.*, 2006). The growing plants were genotyped using a PCR-based procedure that detects the presence of the transposon insertion in the target gene. The PCR products were verified by gel electrophoresis (the primers and PCR conditions are detailed in Table 2.1 and Table 2.2, respectively). The result showed that all of the *PHE1* plants examined were homozygous (Figure 4.2a).

In order to confirm the position of the Ds insertion within the *PHE1* gene the PCR product derived from the gene was verified by DNA sequencing. The sequence was viewed and analysed and the Ds insertion position was compared with the original data in CSHL website (<http://genetrap.cshl.org/>). The *phe1.1* homozygous plants were confirmed to have the Ds insertion at the expected position as detailed on the CSHL website (Figure 4.2b).

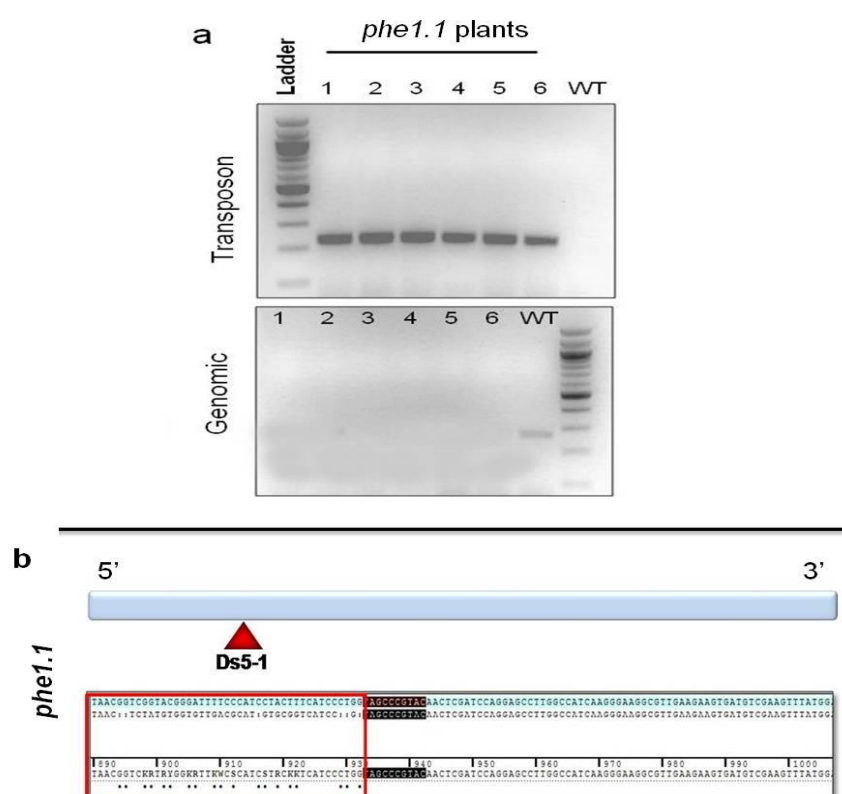


Figure 4.2 Verifying the genotype of *phe1.1* plants. (a) Each lane represents a PCR product of DNA extracted from *phe1.1* plants grown from different seeds obtained from Claudia Kohler, ETH Zurich. Homozygous plants showed bands only in the Ds transposon PCR; the wild type (WT, *Ler*) showed bands only in genomic PCR. The DNA ladder was 100 bp. (b) Schematic showing the Ds transposon insertion site in *phe1.1* plants. The light blue bar represents the exon; the red triangle marks the transposon insertion site. Sequencing confirmed the position of Ds insertion in the *phe1.1* mutant plant, which is highlighted by a red box.

4.2.2 Seed size analyses of *phe1.1* plants

4.2.2.1 Embryo sac area measurement using pollination time and embryo stage

This study focused on endosperm development and its correlation with final seed size. In order to test the role of *PHE1* in endosperm development, selfed seeds of *phe1.1* homozygous plants and wild type (*Ler*) plants, were collected at 0, 3, 4, 5, 6, 7 and 8 day after pollination (DAP) and cleared for microscopy. The embryo sac area of the photographed seeds was measured and the data analysed statistically by either the Student's *t*-test or the Mann-Whitney *U*-test at a 95% confidence level.

Figure 4.3, shows mature ovules and developing seed from the *phe1.1* and the wild type (*Ler*) plants. The *phe1.1* mature ovules and developing seed appeared smaller than wild type. However, seed morphology and the progression of embryogenesis appeared normal and events within both the *phe1.1* and the wild type seeds occurred at approximately the same time e.g. globular embryo at 3 DAP, heart at 5 DAP and mature cotyledons at 8 DAP.

The embryo sac area of the *phe1.1* developing seeds was smaller than the wild type at all time points tested, ranging from a reduction of a 7% ($p = 0.041$) at 5 DAP to a 34% ($p < 0.001$) at 8 DAP (Figure 4.4a). However, the *phe1.1* and the wild type seed displayed very little variation in the stage of embryogenesis relative to DAP. Since the stage of embryogenesis was consistently correlated with DAP, embryo sac area measurements were analyzed at different embryo stages of development for both the *phe1.1* homozygous and the wild type (*Ler*) plants. Developing seeds were taken at the globular, heart, torpedo and mature cotyledons stages of embryogenesis and the embryo sac areas were measured. The embryo sac areas measure of the *phe1.1* was significantly smaller than the wild type at the same development stage, ranging from 9% ($p = 0.010$) at globular embryo stage to 38% ($p < 0.001$) at mature cotyledons stage (Figure 4.4b). These measurements were consistent with the data obtained from the embryo sac areas taken at a different time points (DAPs), showing that the embryo sac area of *phe1.1* seed is small compared to wild type for any particularly stage of embryogenesis or DAP.

In order to examine whether *PHE1* plays a role in the very earliest stages of seed development we decided to examine the effect of *PHE1* on mature ovule size (0 DAP – before fertilization). The mature ovules from *phe1.1* and the wild type plants were collected, cleared and their maximum area measured. The area of mature *phe1.1* ovules

was 23% ($p < 0.001$) smaller than wild type ovules (Figure 4.4b). This observation indicated that *phe1.1* allele showed an early effect on seed development by reducing the size of the mature ovule, which consisted of the maternal tissues only.

In summary, our findings showed that the *phe1.1* allele was associated with a significant reduction in ovule and embryo sac size, suggesting the gene plays an important role in endosperm and seed development as a growth promoter.

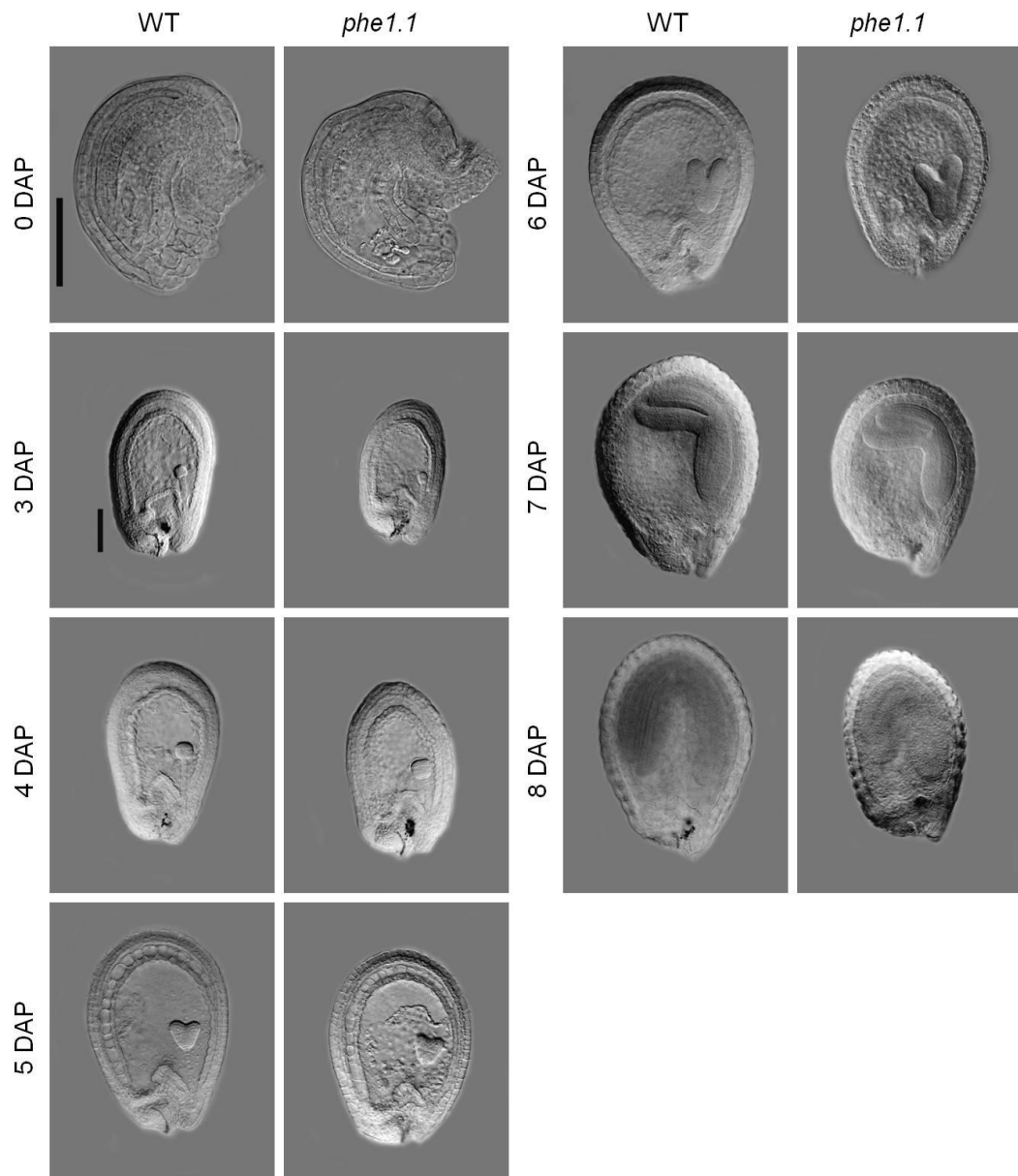


Figure 4.3 Seed development in wild type and *phe1.1* homozygous plants. The cleared mature ovules and seeds were imaged using differential interference contrast (DIC) optics. The wild type is (*Ler*). At 0 DAP (Mature ovule) bar size = 50 μm . At (3 – 8 DAP) bar size = 100 μm . Embryo stage at 3 and 4 DAP is globular; 5 DAP is heart; 6 DAP is late heart; 7 DAP is walking stick; 8 DAP mature cotyledons.

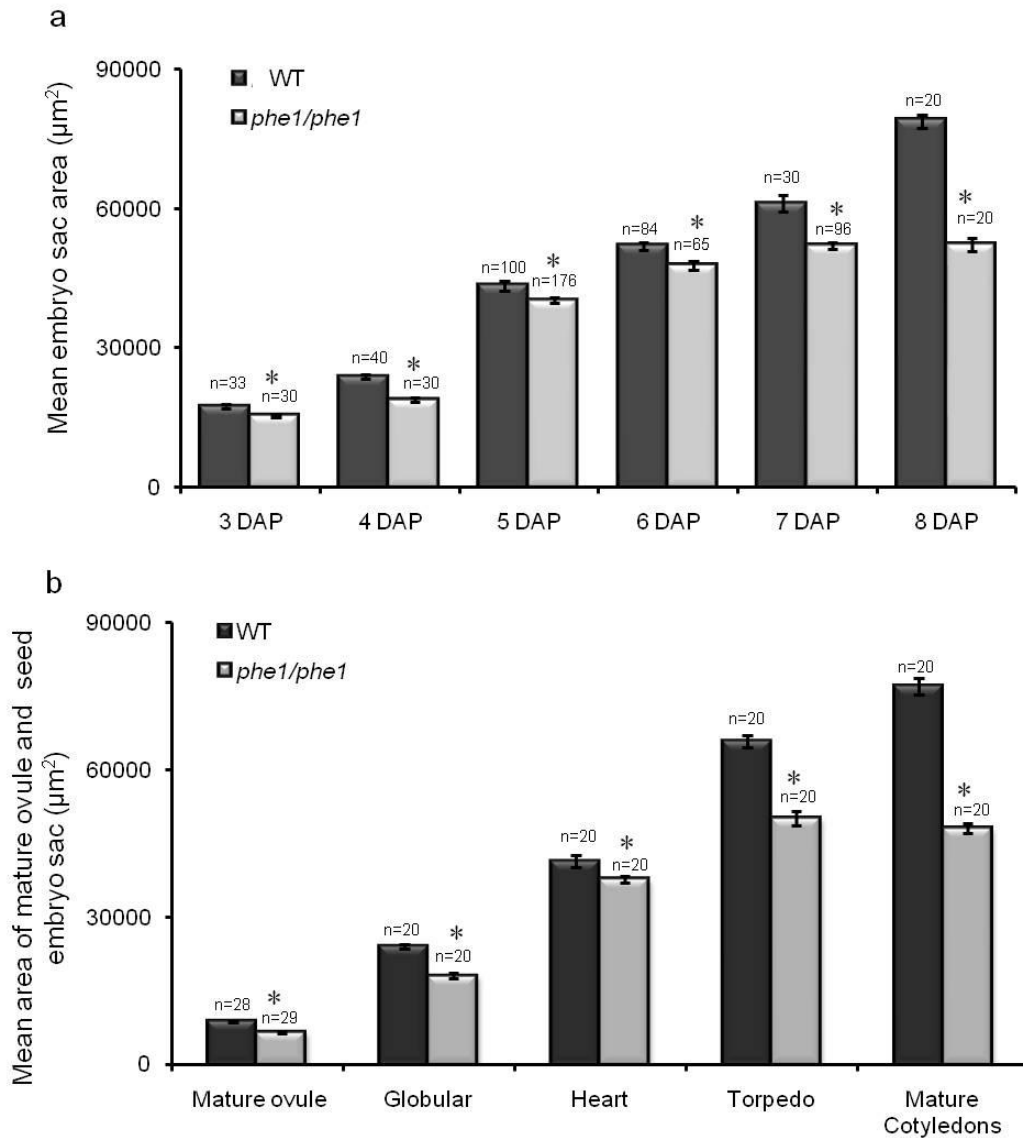


Figure 4.4 Mature ovule and seed phenotype analysis of wild type and *phe1.1* plants. (a) Embryo sac areas at different time points. Columns represent the mean embryo sac areas of the wild type (*Ler*) and the *phe1.1* homozygous at 3, 4, 5, 6, 7 and 8 DAP with a 12 (p=0.001), 21 (p < 0.001), 7 (p = 0.041), 8 (p = 0.007), 15 (p < 0.001) and 34% (p < 0.001) respectively, decrease in area size of *phe1.1* compared to wild type. (b) Mature ovule and embryo sac areas for seed at specific stages of embryogenesis. Columns represent the mean embryo sac areas of the wild type and *phe1.1* homozygotes at mature ovule, globular, heart, torpedo and mature cotyledons with a 24 (p < 0.001), 25 (p < 0.001), 9 (p = 0.010), 24 (p < 0.001) and 38% (p < 0.001) respectively, decrease in size of *phe1.1* compared to wild type. The Student's *t*-test was used and for comparisons, values which differed at the 0.05 significance level between the wild type and the homozygous mutant are labelled with *. Error bar = S.E.M.; n = number of seeds examined

4.2.2.2 Individual seed weight and total yield analyses

The embryo sac area analyses described above indicated that the plants homozygous for the *phe1.1* allele had a significantly smaller embryo sac area size compared to wild type. To determine whether the difference in embryo sac area translated into a reduction in seed size, both individual seed weight and total seed yield were measured for *phe1.1* and wild type plants.

Individual seed weight of the *phe1.1* and its wild type was measured using seeds from ~40 siliques obtained from ten different plants grown under the same conditions. Seeds harvested from each set of siliques were combined (Figure 4.5), weighed in groups of 50 seed, and the mean individual seed weight calculated. The mean individual seed weight of seed from the *phe1.1* homozygous plants was significantly lighter (15.9 μ g; $p = 0.002$) than seed from wild type plants (Table 4.1). Since the level of fertility in terms of number of seed per pod can influence individual seed weight this was examined in a further set of *phe1.1* and wild type plants. Five plants of each of the wild type and the *phe1.1* were grown under the same conditions, five mature pods harvested, and the seeds in each silique counted and the data averaged to determine the seed number/silique. Table 4.1, shows that both the *phe1.1* and the wild type plants had a similar seed number/silique (~43 seeds) indicating that the *phe1.1* allele does not affect seed fertility. This suggests that the reduced seed weight associated with *phe1.1* results from a seed-based effect and was not the result of a trade-off between seed number and size.

In order to determine the effect of the smaller seed size of *phe1.1* plants on the total seed yield five plants of each of *phe1.1* and wild type (*Ler*) were grown under the same conditions and the total seed yield harvested using the Arasystem (illustrated in Chapter 2; Section 2.2.6). The total seeds were harvested after maturation from each plant separately and the total seed weight was calculated separately then averaged. There was significant variation in total seed weight within the set of five plants (Table 4.1). However, the mean total seed yield per plant for *phe1.1* was significantly lower (0.73 g; $p = 0.022$) than the wild type (Table 4.1). Since *phe1.1* and the wild type had almost the same seed number/silique, this suggests that the smaller individual seed size of *phe1.1* had a negative impact on total seed yield. Again this supports the proposition that *PHE1* encodes a growth promoter required during in seed development.

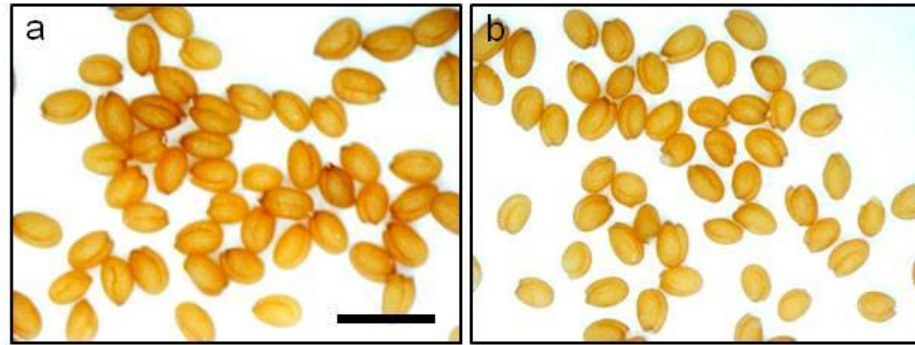


Figure 4.5 Seed phenotype of *phe1.1* and wild type. Photographs of mature dry seeds; the wild type (*Ler*) (a) and *phe1.1* (b). Bar size = 1 mm

Table 4.1 Individual seed weight and total seed yield of *phe1.1* and wild type.

Genotype	Seed weight (μg) \pm S.E.M ^a	Number of seed	Mean number of seed/pod \pm S.E.M ^b	p-value
<i>Ler</i>	18.2 ± 0.49	2192	43 ± 3.6	
<i>phe1.1</i>	15.9 ± 0.25	2187	42 ± 3.7	0.002*
	Total seed yield (g) \pm S.E.M	Number of plant		p-value
<i>Ler</i>	0.88 ± 0.05	5		
<i>phe1.1</i>	0.73 ± 0.02	5		0.022*

^a S.E.M seed weight of $n \geq 36$ siliques; Student's *t*-test with at the 0.05 significant difference level.

^b S.E.M seed set per pod ($n = 5$ pods/ 5 plants); Student's *t*-test with at the 0.05 significant difference level.

* Significantly different compared to wild type.

4.2.2.2.1 Does the size difference between *phe1.1* and wild type seed enable visual identification?

The data above suggests that the reduced individual seed mass of seed produced by selfed *phe1.1* plants resulted from loss of *PHE1* expression. To test the validity of this finding further, an experiment was designed to determine whether *phe1.1/phe1.1* seed could be separated by eye from *phe1.1/PHE1* seed within a mixture of the two genotypes produced in the same pod. The experimental design is illustrated in Figure 4.6. Under a dissecting microscope *phe1.1* emasculated flowers were pollinated by an approximately equal number of pollen from a *phe1.1/phe1.1* plant and wild type (*PHE1*) pollen. Mature siliques were opened under a dissecting microscope and the dry seeds were separated by eye into two size groups, 'large' and 'small'. The mean seed weight was measured before a

proportion of seed from the two groups was sown into soil. Individuals from the resulting seedlings (2-3 weeks old) were selected at random and genotyped. Two control crosses were also generated: one was between a *phe1.1* flower and wild type pollen to create heterozygous seeds with a genotype *phe1.1/PHE1* and the other was a manual selfing of *phe1.1* flowers to produce homozygous seeds with a genotype *phe1.1/phe1.1*.

All of the dry siliques were collected and the total seeds (208) were harvested and then separated into two groups ('large' and 'small') 'by eye' using a low power dissecting microscope. The proportion of the separated large seeds was 59% from the total seed number, while the proportion of the separated small seeds was 41%.

In addition, the two seed size classes and also the seeds from the control crosses were subjected for seed weight analyses. Seeds produced from the manually selfed *phe1.1* plants (*phe1.1 X phe1.1*) and the *phe1.1 X PHE1* cross had a mean seed weight of 15.5 ± 0.2 μg , and (18.6 ± 0.3 μg) respectively. Both control crosses were significantly different in seed weight ($p < 0.001$) (Figure 4.7a).

Figure 4.7a also shows that the mean seed weight of seeds in the 'large' group was significantly lower (16.3 ± 0.9 μg ; $p = 0.017$) compared to the control cross *phe1.1/PHE1* and similar in seed weight to the control cross *phe1.1/phe1.1* ($p = 0.237$). The seeds obtained from the 'small group' were also significantly lighter (15.3 ± 0.3 μg ; $p = 0.001$) than (*phe1.1/PHE1*) seeds and similar to *phe1.1/phe1.1* seeds ($p = 0.574$).

To test whether genotype of the seeds in the two size classes match expectation, 30 seedlings from each seed size class were selected at random and subjected to genotyping. Figure 4.7b shows that approximately half of the total genotyped seedlings belonging to the 'large seed' group were heterozygotes (*phe1.1/PHE1*) and the other half were homozygotes (*phe1.1/phe1.1*). In contrast, the majority of seed ($79.0\% \pm 0.9$) within the 'small seed' group yielded *phe1.1/phe1.1* seedlings. This data shows that the 'small' seed class was as expected enriched for the *phe1.1/phe1.1* genotype indicating that the impact of the loss of *PHE1* expression on seed size has sufficiently large impact on seed size to enable selection of the seed with this genotype by eye. However, the 'large' seed class did not produce the expected result indicating that there is some overlap in the size of *PHE1/phe1.1* and *phe1.1/phe1.1* seed.

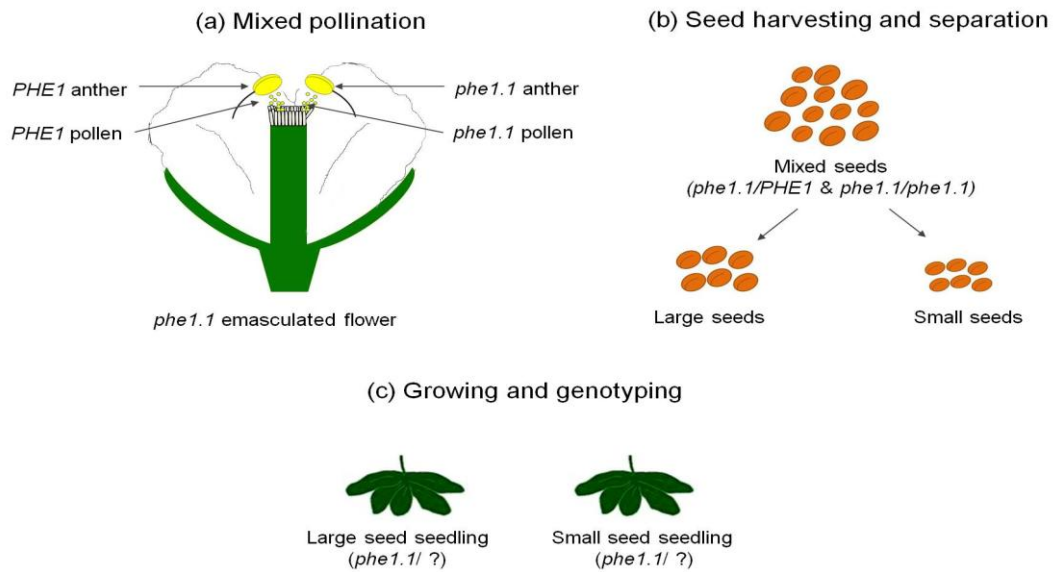


Figure 4.6 Experimental design of visual seed size selection experiment (a) Mixed pollination of *phe1.1* emasculated flower with equal amount of *phe1.1* and wild type (*PHE1*) pollen. (b) Mature dry seed harvesting and separation into two size classes by eye (c) Growth and genotyping of seedlings from the two seed classes.

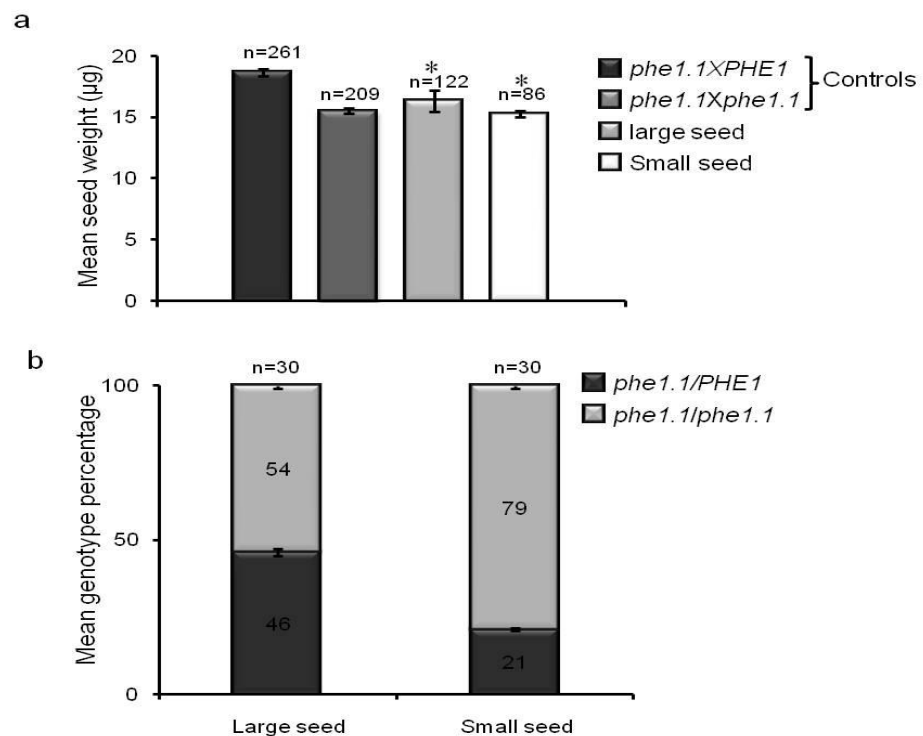


Figure 4.7 Weight and genotypes of seed derived from *phe1.1* crosses with *phe1.1* and wild type. (a) Columns represent the mean seed weight of the control crosses, 'large' seed and 'small' seed. (b) Columns represent the proportion of heterozygous (*phe1.1/PHE1*) and homozygous (*phe1.1/phe1.1*) genotypes determined from seedlings grown from each seed group ('large' and 'small'). For comparison, values that differ at the 0.05 significance level from one or both controls are labelled with *. Error bar = S.E.M.; n = number of seeds examined.

4.2.2.3 Relationship between embryo sac area, seed mass and endosperm size in *phe1.1* and wild type

Since *phe1.1* reduces both embryo sac area and the mass of mature seed, we sought to directly determine whether loss of *PHE1* results in a reduction in the size of the endosperm. The most straight-forward method to quantify the size of the endosperm in developing seeds is to count the number of peripheral endosperm nuclei using confocal microscopy (Scott *et al.*, 1998). Previous studies have shown that seed at 4 and 5 DAP are optimal, since at these stages the endosperm is ending its proliferative phase and beginning to differentiate as evidenced by cellularisation. Changes to endosperm development brought about by, for example, genomic imbalance are thus obvious either by changes in the number of endosperm nuclei or to the timing of cellularisation.

Confocal slides were prepared of *phe1.1* and wild type (*Ler*) seed at 4 and 5 DAP (late globular to early heart embryo) to count endosperm. The number of endosperm nuclei was obviously less in *phe1.1* seed than the wild type at both 4 DAP and 5 DAP (Figure 4.8a). This was confirmed by endosperm nuclei counts which showed that *phe1.1* seed contained an average of 191 ± 6 nuclei and the wild type contained 336 ± 8 nuclei (Figure 4.8b). This represents a difference of 43% ($p < 0.001$) between the two genotypes.

In summary, our finding shows that reduced seed mass associated with the *phe1.1* genotype is correlated with a dramatic reduction in the number of peripheral endosperm nuclei formed during early seed development. This loss suggests that *PHE1* plays a role in promoting endosperm proliferation.

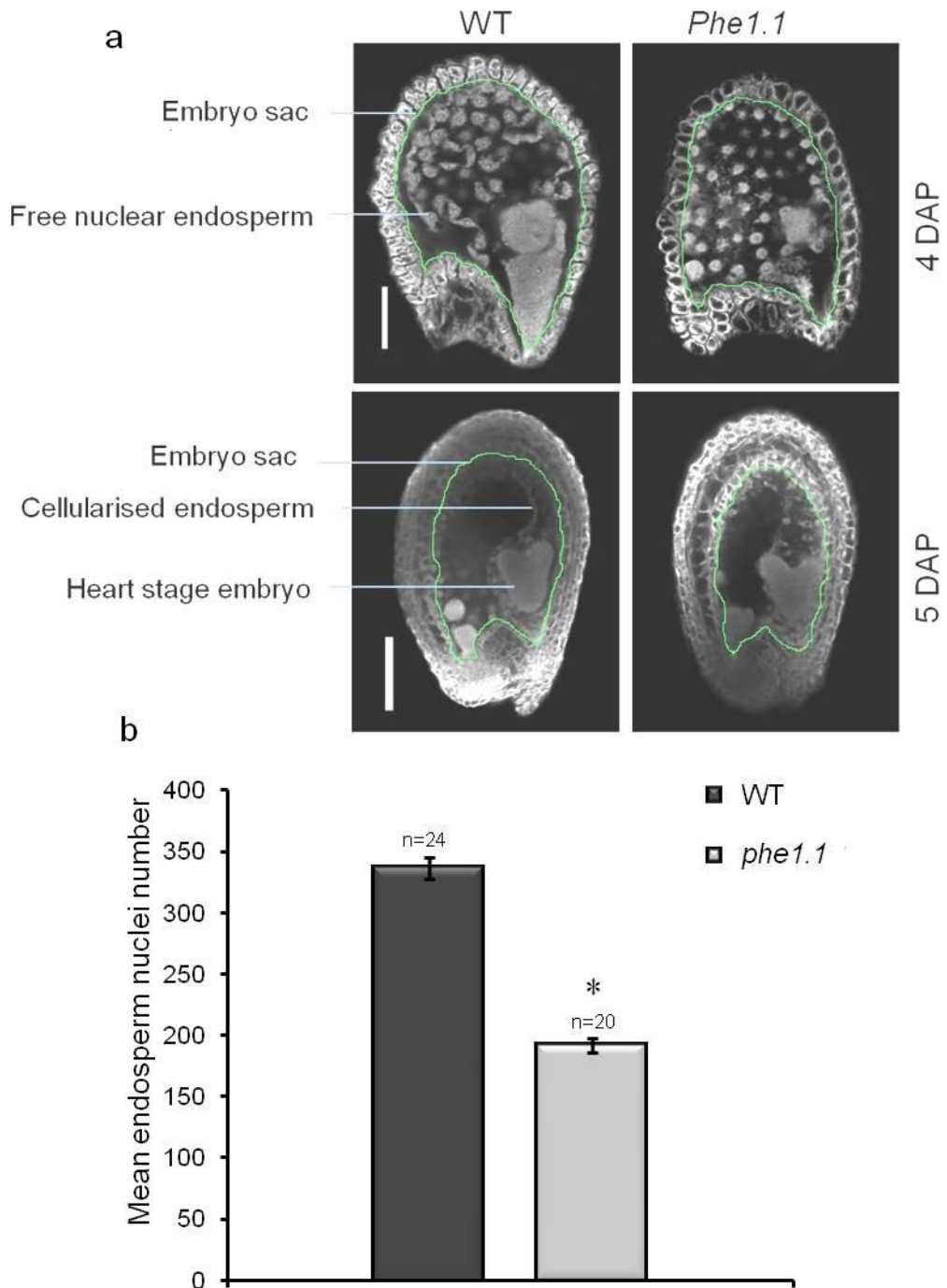


Figure 4.8 Endosperm nuclei number and cellularisation in *phe1.1* and wild type seed. (a) Confocal images of *phe1.1* and wild type (*Ler*) seed showing the embryo sac (highlighted by green) and free nuclear endosperm at 4 DAP; the whole seed phenotype at 5 DAP detailing the cellularisation of the peripheral endosperm; bar size at 4 DAP = 50 μ m and 5 DAP = 100 μ m. (b) Columns represent the mean endosperm nuclei number of *phe1.1* and wild type at 4 and 5 DAP. For comparison, values that differ at the 0.05 significance level from wild type are labelled with *. Error bar = S.E.M.; n = number of seeds examined.

4.2.3 The role of *PHE1* in interploidy crosses

The data presented above strongly suggest that *PHE1* encodes a promoter of endosperm proliferation as predicted by the parental conflict model (Haig and Westoby, 1989, 1991). The current model for the regulation of *PHE1* imprinting suggests that the maternal *PHE1* allele is silenced by the FIS-PcG complex, which includes *MEDEA* (Kohler *et al.*, 2005; Makarevich *et al.*, 2006). Since maternally-inherited loss-of-function alleles of *MEDEA* result in massive endosperm over-proliferation, embryo arrest and seed abortion (Grossniklaus *et al.*, 1998; Kinoshita *et al.*, 1999; Kiyosue *et al.*, 1999; Luo *et al.*, 1999), one possibility is that the resultant ectopic expression of the maternal *PHE1* allele in a *mea* seed (Kohler *et al.*, 2005) is partly or wholly responsible for the endosperm over-growth phenotype.

The 2x X 4x paternal excess interploidy cross also results in endosperm over-proliferation in *A. thaliana* (Scott *et al.*, 1998). Since the endosperm produced by this cross contains an extra paternal genome, and therefore an extra active copy of the *PHE1* gene, the possibility exists that this endosperm over-growth phenotype, like that in *mea* seeds, is also generated by *PHE1* over-expression. If this is true, it follows that seed sired on a wild type diploid seed parent by a tetraploid pollen parent carrying a loss-of-function *PHE1* gene will exhibit a reduced level of endosperm over-proliferation and a proportional reduction in seed size. Conversely, since expression of the maternal *PHE1* allele is relatively low, a loss-of-function *PHE1* allele should have little or no impact on the outcome of a 4x X 2x maternal excess cross. To test these propositions, 4x *phe1.1* plants were generated and used in various diagnostic crosses.

4.2.3.1 Seed weight and abortion frequency analyses of 4x *phe1.1* reciprocal interploidy crosses

Reciprocal interploidy crosses (♀ 4x X ♂ 2x) and (♀ 2x X ♂ 4x) using wild type diploid and *phe1.1* tetraploid plants, as well as a control wild type reciprocal crosses, were made and the resulting seed phenotypes for size, weight and the level of any seed abortion quantified. As expected the pair of control (*PHE1*) interploidy crosses produced seed of the two expected size classes, ♀ 4x X ♂ 2x seed were small and ♀ 2x X ♂ 4x were large (Figure 4.9a). This was also the case for crosses with the 4x *phe1.1* genotype, suggesting that loss of *PHE1* function did not, as proposed above, reduce the size of seed produced in a ♀ 2x X ♂ 4x cross.

Confocal images of a developing seeds were produced from a normal diploid (2x X 2x) and interploidy (4x X 2x, 2x X 4x) crosses at 6 DAP to compare the general seed phenotype (embryo stage and timing of peripheral endosperm cellularisation) in the 4x *phe1.1* and wild type crosses. As expected, there was a significant difference in embryo sac area and stage of peripheral endosperm cellularisation between the normal diploid cross (2x X 2x) and interploidy crosses (4x X 2x and 2x X 4x) as reported by Scott *et al.* (1998). However, in accord with the findings above relating to seed size there were no discernable differences in general seed phenotype and endosperm cellularisation time in seed generated by the 4x *phe1.1* and wild type interploidy crosses (Figure 4.9b-g).

Next, seed size in the various crosses was subjected to quantitative analysis. Individual seed size was determined by weighing all seeds present in each silique (plump and shrivelled). In the paternal excess (2x X 4x) set of crosses, the wild type 2x *PHE1* X 4x *PHE1* cross produced seed that was slightly smaller ($27.0 \mu\text{g} \pm 0.5$; $p < 0.001$) than the 2x *PHE1* X 4x *phe1* cross ($29.2 \mu\text{g} \pm 0.4$) (Figure 4.10a). This confirmed that loss of *PHE1* function in a tetraploid pollen parent does reduce the relative overgrowth in seed mass associated with a 2x X 4x cross. In contrast, the maternal excess crosses (4x X 2x) yielded an unexpected outcome: seed weight was reduced in the 4x *phe1* X 2x *PHE1* cross ($9.0 \mu\text{g} \pm 0.3$) relatively to the control 4x *PHE1* X 2x *PHE1* ($14.4 \mu\text{g} \pm 0.3$; $p < 0.001$) cross. This indicated that loss of *PHE1* function resulted in a reduction in seed size, suggesting that *PHE1* may play a growth-promoting role within the endosperm when expressed from maternal genome.

The seed weight data described above represented the individual seed weights derived by averaging seed weight for all the seed, both plump and shrivelled, collected from individual siliques. Since the mass of plump and shrivelled seeds differed markedly, the relative levels of seed abortion in the various crosses will have influenced the individual seed weight. Figure 4.10b, shows the mean seed abortion frequency in the siliques used in the seed weight analysis. The 4x *PHE1* X 2x *PHE1* wild type cross, which had a relatively high seed weight, showed a significantly lower seed abortion frequency (6%; $p < 0.001$) compared to the 4x *phe1* X 2x *PHE1* (16%) cross. This indicates that the individual seed weight in this cross was reduced due to a relatively high level of seed abortion, and that loss of maternal *PHE1* function reduces seed viability in 4x X 2x crosses. In contrast, the abortion frequency in the 2x X 4x crosses exhibited the opposite trend, with the 2x *PHE1* X 4x *PHE1* wild type cross showing a higher seed abortion (13%; $p < 0.001$) than the 2x

PHE1 X 4x *phe1* cross (5%). This indicates that loss of paternal *PHE1* function dramatically reduced seed lethality associated with a 2x X 4x cross.

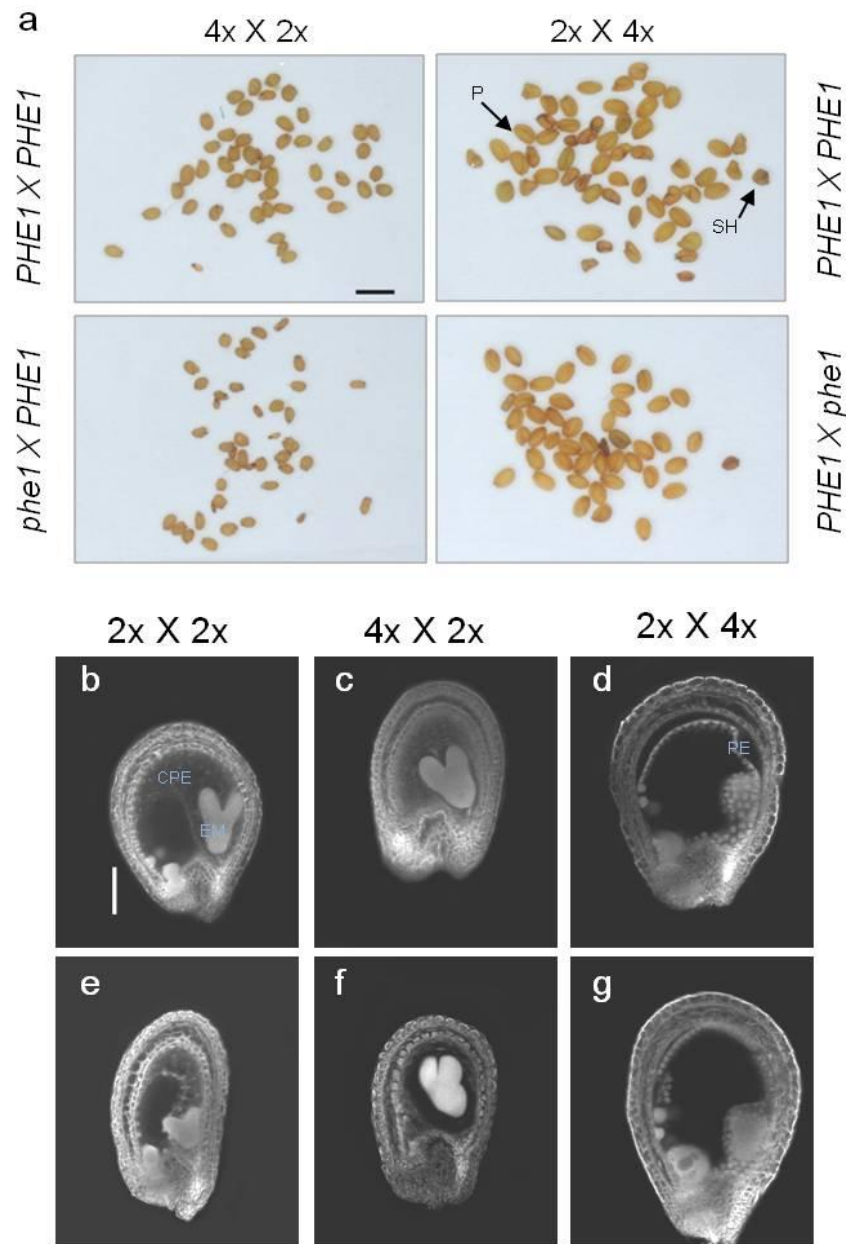


Figure 4.9 The effect of *PHE1* loss of function on the outcome of interploidy crosses.

(a) Mature dry seeds of 4x X 2x and 2x X 4x crosses representing wild type crosses at the top and 4x *phe1* crosses at the bottom also the images show the size and phenotype of plump (P) and shrivelled (SH) seed on each cross; bar size = 1mm. Confocal images of 2x X 2x, 4x X 2x and 2x X 4x crosses of a wild type (*Ler*) and *phe1* mutant at 6 days after pollination (DAP); (b) 2x X 2x wild type seed with a heart embryo (EM) stage and cellularised peripheral endosperm (CPE); (c) 4x X 2x wild type seed with heart stage embryo and completely cellularised endosperm; (d) 2x X 4x wild type seed with heart stage embryo and the peripheral endosperm (PE) started to cellularise; (e) 2x X 2x *phe1* mutant seed with a heart embryo and cellularised endosperm; (f) 4x *phe1* X 2x *PHE1* seed with heart embryo and completely cellularised endosperm; (g) 2x *PHE1* X 4x *phe1* seed with heart stage embryo and the endosperm just started to cellularise. Bar size = 100 μ m.

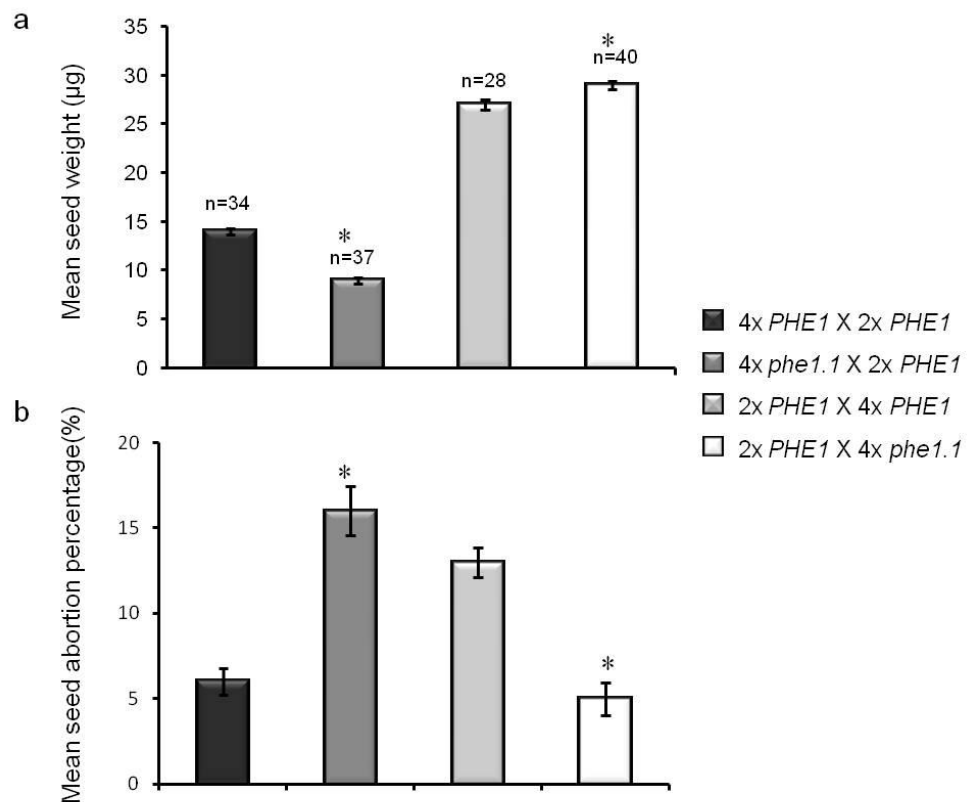


Figure 4.10 Seed phenotype analysis of 4x *phe1.1* and wild type interploidy crosses. (a) Columns represent the mean seed weight of (4x X 2x) and (2x X 4x) crosses for 4x *phe1.1* and wild type (*Ler*). (b) Columns indicate the mean seed abortion in percentage (%) for the same crosses illustrated in (a). For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *. Error bar = S.E.M.; n = number of siliques examined.

4.2.3.2 Seed weight and abortion frequency analyses of a restricted (4x) *phe1.1* reciprocal interploidy crosses

To confirm the result observed in a previous section that using 4x *phe1.1* in the paternal excess cross lowered seed abortion and to exclude any physiological factors which may affect the result such as the silique position on the plant, a restricted pollination experiment was carried. In restricted pollination, only 3 flowers on the main inflorescence were selected for emasculation and crossing, and all the other branches on the plant were left to self-pollinate. In the restricted pollination the mean seed weight of the wild type cross (4x *PHE1* X 2x *PHE1*) was significantly higher ($13.7 \mu\text{g} \pm 0.3$) than the 4x *phe1.1* X 2x *PHE1*) cross ($9.0 \mu\text{g} \pm 0.6$; $p < 0.001$), whereas the 2x *PHE1* X 4x *phe1.1* cross ($32.0 \mu\text{g} \pm 0.4$; $p = 0.140$) showed no significant different in seed weight compared to the wild type cross (2x *PHE1* X 4x *PHE1*; $30.7 \mu\text{g} \pm 0.8$) (Figure 4.11a).

Although the 4x *phe1.1* X 2x *PHE1* cross showed a high seed weight compared to wild type cross, the mean seed abortion frequency of the same siliques were not significantly different: 16% \pm 2.4 vs. 13% \pm 1.4 ($p = 0.366$). In contrast, the seed abortion frequency of 2x *PHE1* X 4x *phe1.1* cross was significantly lower (57%; $p = 0.002$) compared to the wild type cross (Figure 4.11b). The seed weight and abortion frequency data obtained by restricted pollination were in close agreement with the data obtained from unrestricted pollination described in the previous section.

In summary, our findings in both unrestricted and restricted pollination experiments have suggested that loss of *PHE1* gene function results in changes to the frequency of seed abortion in interploidy crosses. Maternal loss increases abortion in the 4x X 2x cross, whilst paternal loss greatly reduces seed abortion in the 2x X 4x cross.

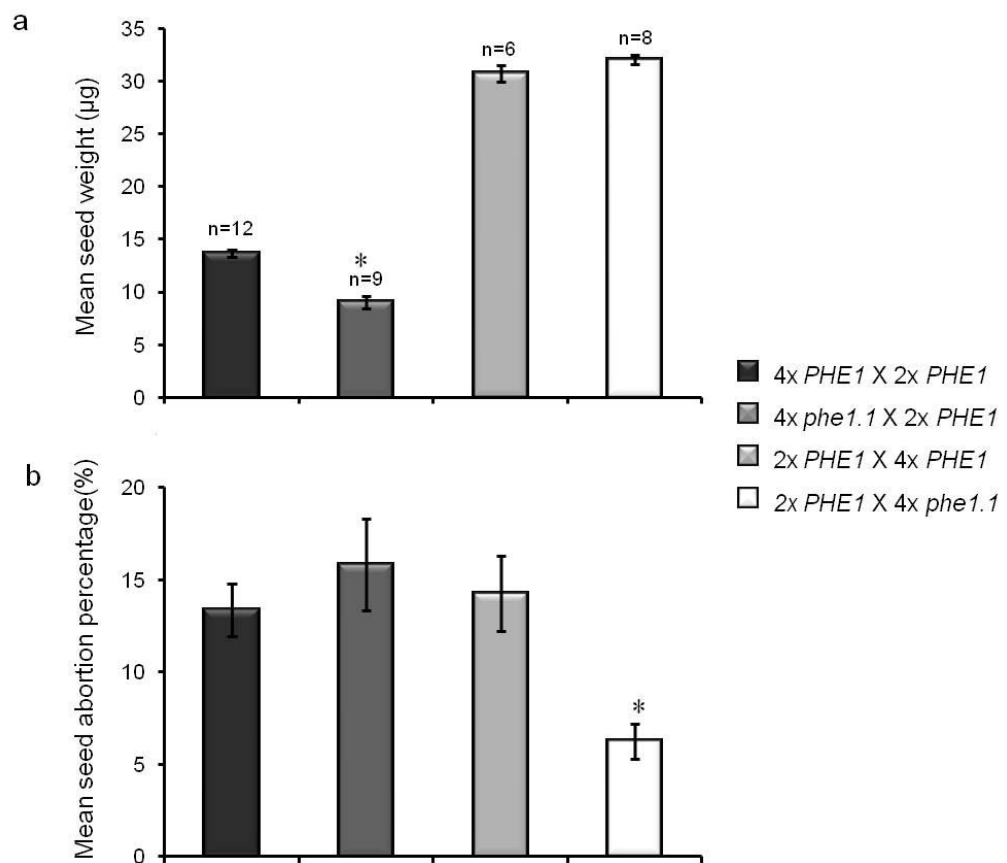


Figure 4.11 Seed phenotype analysis of 4x *phe1.1* and wild type interploidy crosses produced from restricted pollination. (a) Columns represent the mean total seed weight of the (4x X 2x) and (2x X 4x) crosses for 4x *phe1.1* and the wild type (*Ler*). (b) Columns stand for the mean seed abortion in percentage (%) for the same crosses illustrated in (a). For comparisons, values that differ at the 0.05 significance level from wild type are labelled with *. Error bar = S.E.M.; n = number of siliques examined.

4.3 Discussion

The primary aim of the work reported in this chapter was to examine the role of *PHE1* in seed size regulation in *A. thaliana*. *PHE1* is an imprinted gene that is expressed paternally and suppressed maternally (Kohler *et al.*, 2005); therefore according to the parental conflict model (Haig and Westoby, 1989; 1991) *PHE1* should encode growth promoter. The work in this chapter illustrated the first study of *PHE1* loss-of function mutant phenotype.

4.3.1. *PHE1* loss-of-function overall vegetative phenotype appear like wild type

Overall plant vegetative phenotype such as stem length, leaf shape and size, number of branches, flower symmetry, and flowering time of *phe1.1* mutant was similar to wild type consistent with previous findings (Kohler *et al.* 2005) and expression profile data (Kohler *et al.*, 2003a; Day *et al.*, 2008; Tiwari *et al.*, 2010; see also Table 3.2).

4.3.2 *PHE1* loss-of function seed phenotype

4.3.2.1 The *phe1.1* allele shows a reduction in endosperm, individual seed size and total seed yield

Analysis of embryo sac areas of *phe1.1* and wild type (*Ler*) indicated that *phe1.1* has a significantly smaller embryo sac area than wild type at all the time points examined (3-8 DAP). Embryo sac areas measured at different embryo developmental stages provided essentially the same result (Figure 4.4a). The mature ovules of *phe1.1* plants were 23% ($p < 0.001$) smaller than wild type ovules (Figure 4.4b). This observation indicated that *phe1.1* allele showed an early effect on seed development. These findings suggest that *phe1.1* allele is associated with a significant reduction in ovule and embryo sac size, suggesting it plays a significant role in endosperm and seed development as a growth promoter.

Seed weight analysis revealed that the *phe1.1* plants produced significantly smaller than wild type plants with very similar numbers of seed per pod. This reduced individual seed size appeared to have a strong negative impact on total seed yield (Table 4.1). This data expands on that described by Josefsson *et al.* (2006) who reported that the *phe1.1* seeds were viable and somewhat smaller than wild type, but did not provide any in depth quantitative analysis.

4.3.2.1.1 The small seed size of *phe1.1* is apparent compared to wild type

We suggested in the previous section that the reduced mass of seed produced by selfed *phe1.1* plants resulted from loss of *PHE1* expression. In order to test the validity of this finding further, an experiment was designed (Figure 4.7) to determine whether *phe1.1/phe1.1* seed could be separated by eye from *phe1.1/PHE1* seed within a mixture of the two genotypes produced in the same pod. We were successful in confirming the small size of *phe1.1* by endosperm and seed size analysis (Table 4.1; Figure 4.4).

We also demonstrated that *phe1.1* seeds could be selected according to their seed size from a mixture of seeds having *phe1.1/phe1.1* and *phe1.1/PHE1* genotypes presented in the same pod. The seeds obtained from the ‘large seed’ group were significantly smaller ($16.3 \pm 0.9 \mu\text{g}$; $p = 0.017$) than (*phe1.1/PHE1*) and similar to (*phe1.1/phe1.1*) seeds. Also, the selected ‘large seed’ group showed a 50% of seeds with an expected genotype (*phe1.1/PHE1.1*). In contrast, the seeds obtained from the ‘small seed’ group were significantly smaller ($15.3 \pm 0.3 \mu\text{g}$; $p = 0.001$) than (*phe1.1/PHE1*) seed and similar to (*phe1.1/phe1.1*) seed in weight. Significantly, 79% of the selected ‘small seed’ group had the expected *phe1.1/phe1.1* genotype indicating that *phe1.1* does indeed significantly reduce seed size. These findings on embryo sac area, seed weight, total yield and the apparent small seed size of *PHE1* loss-of-function phenotype suggest that *PHE1* has a significant role in endosperm proliferation and development which translated into an effect on final seed size.

4.3.2.2 The reduction in endosperm nuclei number is associated with the reduction in seed mass and endosperm size of *phe1.1*

Since *phe1.1* reduces both embryo sac area and the mass of mature seed, we attempted to directly determine whether loss of *PHE1* resulted in a reduction in the size of the endosperm. The *phe1.1* endosperm had a significantly less endosperm free nuclei than wild type endosperm and both *phe1.1* and wild type endosperm was cellularised at the same time (5 DAP) (Figure 4.8a). Since, there was no difference between *phe1.1* and wild type in the timing of endosperm cellularisation, the reduced endosperm nuclei number in *phe1.1* most likely accounts for the reduced seed size with loss of *PHE1* function.

4.3.3 Extra copy of paternal *PHE1* potentially contributes to seed death in interploidy crosses

PHE1 is an imprinted gene in *A. thaliana* that is preferentially expressed from the paternal allele during early endosperm development (Kohler *et al.*, 2005). Since the loss-of-function of *PHE1* gene reduces endosperm proliferation resulting in small seeds the *PHE1* gene most likely encodes a promoter of endosperm proliferation as predicted by the parental conflict model (Haig and Westoby, 1989; 1991).

The current model for the regulation of *PHE1* imprinting suggests that the maternal *PHE1* allele is silenced by the FIS-PcG complex, which includes *MEDEA* (Kohler *et al.*, 2005; Makarevich *et al.*, 2006). Since maternally-inherited loss-of-function alleles of *MEDEA* result in massive endosperm over-proliferation, embryo arrest and seed abortion (Grossniklaus *et al.*, 1998; Kinoshita *et al.*, 1999; Kiyosue *et al.*, 1999; Luo *et al.*, 1999), one possibility is that the resultant ectopic expression of the maternal *PHE1* allele in a *mea* seed (Kohler *et al.*, 2005) is partly or wholly responsible for the endosperm overgrowth phenotype. Since 2x X 4x interploidy crosses produce similar effects on endosperm proliferation (Scott *et al.*, 1998) we used a 4x *phe1.1* line in reciprocal interploidy crosses to investigate whether the endosperm over-growth phenotype is also generated by *PHE1* over-expression. If true, it was reasoned that seed sired on a wild type diploid seed parent by a tetraploid pollen parent carrying a loss-of-function *PHE1* gene would have exhibited a reduced level of endosperm over-proliferation and a proportional reduction in seed size overgrowth. Conversely, since expression of the maternal *PHE1* allele is relatively low, a loss-of-function *PHE1* allele was expected to have little or no impact on seed size in a 4x X 2x cross.

In terms of seed size, as measured by seed weight in unrestricted and restricted crosses, the outcome of the pair of 2x X 4x crosses apparently did not support a significant role for *PHE1* in the seed overgrowth phenotype. In fact, rather than reducing seed size, the 2x X 4x *phe1.1* cross resulted in seed with a higher individual seed weight than the control cross (Figure 4.10). However, this increase in individual seed weight was in part due to a dramatic reduction (5% vs. 13%; $p < 0.001$) in the frequency of shrivelled seed produced in the 2x X 4x *phe1.1* compared to the control cross. This indicates that lethal endosperm over-proliferation that causes seed abortion in interploidy crosses in *A. thaliana* (Scott *et al.*, 1998; Adams *et al.*, 2000; Comai *et al.*, 2000; Bushell *et al.*, 2003; Dilkes and Comai, 2004; Dilkes *et al.*, 2008) requires paternal *PHE1* expression.

Interestingly, our data point to a maternal role of *PHE1*. Individual seed weight for the 4x *phe1.1* X 2x cross was lower than the control 4x X 2x cross, indicating that loss of maternal *PHE1* function reduced seed size. However, as with the 2x X 4x crosses, individual seed weight measurements masked a strong effect on seed abortion frequency. In this case, the 4x *phe1.1* X 2x cross resulted in a dramatic increase in the frequency of shrivelled seed (16% vs. 6%; $p < 0.001$) over the control cross. This effect was unexpected since *PHE1* is preferentially expressed from the paternal genome (Kohler *et al.*, 2005) consistent with its role as an imprinted growth promoter (Haig and Westoby, 1991). Clearly, however, maternal expression of *PHE1* plays a role in seed development perhaps by providing PHE1 protein during the very earliest stages of endosperm development when the expression from the paternal *PHE1* allele may be limited (Vielle-Calzada *et al.*, 2000). Lack of *PHE1* expression during this period may therefore reduce endosperm size below a critical level in a proportion of seed resulting in increased seed abortion.

The data described above supports previous work on exploring the role of *PHE1* in the relationship between hybrid incompatibility and gene silencing in intra- and -interploidy crosses involving *A. thaliana* and *A. arenosa* (Josefsson *et al.*, 2006). In interspecific interploidy crosses between *A. thaliana* and *A. arenosa*, where a 2x *A. thaliana* seed parent is crossed with a 4x *A. arenosa* pollen parent, do not produce viable seeds. However, viable seeds are produced when a 4x *A. thaliana* seed parent is crossed with the 4x *A. arenosa* pollen parent (Comai *et al.*, 2000; Bushell *et al.*, 2003). Bushell *et al.* (2003) suggested that the foundation for the success or failure of these interspecific crosses remains in the endosperm. The authors also demonstrated that endosperm from the 2x *A. thaliana* X 4x *A. arenosa* had phenotypes characteristic of paternal genomic excess, such as a lack of cellularisation and prolonged proliferation.

Josefsson *et al.* (2006) reported that in both intra- and interploidy crosses involving *A. thaliana* and *A. arenosa* as a pollen parent, imprinting of *PHE1* was lost with biallelic expression of paternal and maternal alleles in the endosperm. As mentioned earlier, imprinting of *PHE1* is due to the repressive effects of a maternal PcG complex. The author found that the loss of *PHE1* imprinting was due to an excess of the PcG complex target sites in the *A. arenosa* paternal genome compared to the *A. thaliana* paternal genome. Therefore, the excess of the target sites in the *A. arenosa* paternal genome could overcome the dosage of maternal PcG complexes, allowing the maternal *PHE1* allele to avoid

complete silencing which causes seed abortion (Josefsson *et al.*, 2006), presumably by promoting excessive endosperm over-proliferation.

Josefsson *et al.* (2006) also suggested that over-expression of *PHE1* from the maternal genome contributed to hybrid seed death in *Arabidopsis* crosses and that knocking out maternal *PHE1* rescued hybrid seed survival. Our study in 4x *phe1.1* interploidy crosses supports this idea by showing that *PHE1* has an important role in controlling seed abortion in both 2x X 4x and 4x X 2x cross within *A. thaliana*. Both studies give support to the differential dosage hypothesis which suggests that imprinting developed to control the relative dosage of the regulatory factors in the endosperm (Dilkes and Comai, 2004).

In conclusion, our study in *phe1.1* phenotypic characterisation suggests that *PHE1* plays a critical role in endosperm growth and development as a growth promoter.

4.3.4 Future work

PHE1 is an imprinted gene that is expressed paternally and suppressed maternally by the FIS complex (Section 4.2). *MEA* belongs to the FIS complex which is active in the female gametophyte and in the endosperm and causes suppression of *PHE1* expression before fertilization. In addition, up-regulation of *PHE1* in *mea* mutants is partially responsible for the *mea* mutant phenotype (Kohler *et al.*, 2003a). Therefore, in order to understand the behaviour of *PHE1* in 4x *phe1* X 2x *PHE1* and 2x *PHE1* X 4x *phe1* crosses we should examine *MEA* expression in these and wild type interploidy crosses.

Chapter 5: Identifying the role of *AGL62* in endosperm growth and development

5.1 Introduction

The endosperm is an important part of the seed since it both supplies nutrients to the developing embryo during seed growth and participates in seed germination. The work described in Chapter 4 examined the role of *PHERES 1* (*PHE1*) in endosperm development and reported the in depth study of a *PHE1* loss-of-function mutant phenotype. This chapter focuses on testing the role of another candidate gene, *AGL62*, which was similarly associated with endosperm over-proliferation (Chapter 3). This investigation employed two strategies: knock-out (KO), where the target gene was inactivated by a mean of T-DNA insertion and knock-in (KI) where the target gene was over-expressed from a peripheral endosperm promoter.

5.1.1 *AGL62* is a type 1 MADS box transcription factor associated with endosperm over-proliferation

Crosses between plants of different ploidies alter endosperm growth in *Arabidopsis thaliana* (*A. thaliana*) (Scott *et al.*, 1998). For example, a 4x X 2x cross produced a small seed containing a small embryo and a small, prematurely cellularised, endosperm (maternalized endosperm), while in a 2x X 4x cross the resulting seeds were large, contained a large embryo and the endosperm was large due to increased endosperm proliferation (paternalized endosperm). As described in Chapter 3, microarray-based transcript profiling of RNA isolated from seed generated by various interploidy crosses in *A. thaliana* produced a list of genes associated with endosperm over- or under-proliferation. One such gene was *AGL62*, which showed high expression in seeds that contained a paternalized endosperm and was therefore classified as positively associated with endosperm proliferation with an exclusive expression in the endosperm (Tiwari *et al.*, 2010; Chapter 3; Table 3.2). The *AGL62* expression level was twofold and sevenfold higher in the paternal excess crosses 2x X 4x and 2x X 6x, respectively, compared to 2x X 2x cross (Chapter 3; Table 3.2).

AGL62 belongs to the MADS box gene family of transcription factors (Parenicova *et al.*, 2003). MADS box proteins contain a conserved MADS box domain that functions in DNA binding (Shore and Sharrocks, 1995; West *et al.*, 1997; Parenicova *et al.*, 2003). Although the function of many of these genes is unknown, some have been allocated roles

in different aspects of plant development such as plant reproduction, flowering time, floral organogenesis, floral meristem identity, fruit formation, seed pigmentation and endothelium development (Ng and Yanofsky, 2001; Nesi *et al.*, 2002; Becker and Theissen, 2003). *AGL62* belongs to the type 1 MADS box group which consists of ~61 genes in *A. thaliana* (Alvarez-Buylla *et al.*, 2000a; Parenicova *et al.*, 2003). A few members have been functionally characterised including: *PHE1* (*AGL37*), which imprinted in endosperm and preferentially expressed from the paternal allele (Kohler *et al.*, 2003b); *AGL80* that plays a role in central cell or endosperm development (Portereiko *et al.*, 2006); *AGL28* which causes precocious flowering when over-expressed (Yoo *et al.*, 2006); and *AGL23* that controls female gametophyte and embryo development (Colombo *et al.*, 2008). Yeast two-hybrid technology showed that the *AGL62* protein interacts with other proteins belonging to the type 1 MADS box protein family such as *PHE1*, *PHE2*, *AGL97*, *AGL90*, *AGL36*, *AGL92*, *AGL86* and *AGL80* (de Folter *et al.*, 2005).

Thus, given its expression pattern, *AGL62* may function as a transcriptional regulator of some aspect of endosperm development in the *A. thaliana* seed. This chapter reports experiments aimed at testing this hypothesis using inactivation (KO) and over-expression (KI) of the gene and an assessment of their impact on endosperm and seed development.

5.2 Results

5.2.1 Analysing of *AGL62* loss-of-function phenotype

As explained earlier (Chapter 1; Section 1.7), the KO strategy was intended to cause gene inactivation by mean of a T-DNA insertion that disrupts the DNA of the target gene. If a growth promoter gene that contributes to increased seed size is inactivated this is expected to result in a reduced seed size. Inactivation of a growth inhibitor gene would have the opposite effect on seed size. The exact outcome was expected to depend on whether the gene is endosperm specific, seed specific or expressed generally.

5.2.1.1 Genotyping *AGL62* T-DNA insertion line

In order to verify the genotype of the *AGL62* T-DNA insertion line (SALK-022148; Figure 5.1b) before engaging in further experiments, 40 plants were grown from seed obtained from the Nottingham Arabidopsis Stock Centre (NASC) and genotyped using a PCR-based procedure that detects the presence of the T-DNA in the target gene. The PCR products were verified by gel electrophoresis (the primers and PCR conditions are detailed in Table

2.1 and Table 2.2, respectively). The results showed that all SALK-022148 individuals were heterozygous (Figure 5.1a). The failure to recover homozygotes was investigated further by selfing several *AGL62/agl62* heterozygote plants and analysing a large number (>200) of the resulting offspring for the *agl62/agl62* genotype. Again, no *agl62/agl62* individuals were recovered; plants were either wild type or heterozygous in an approximately 1:2 ratio (74 *AGL62/AGL62*: 156 *AGL62/agl62*).

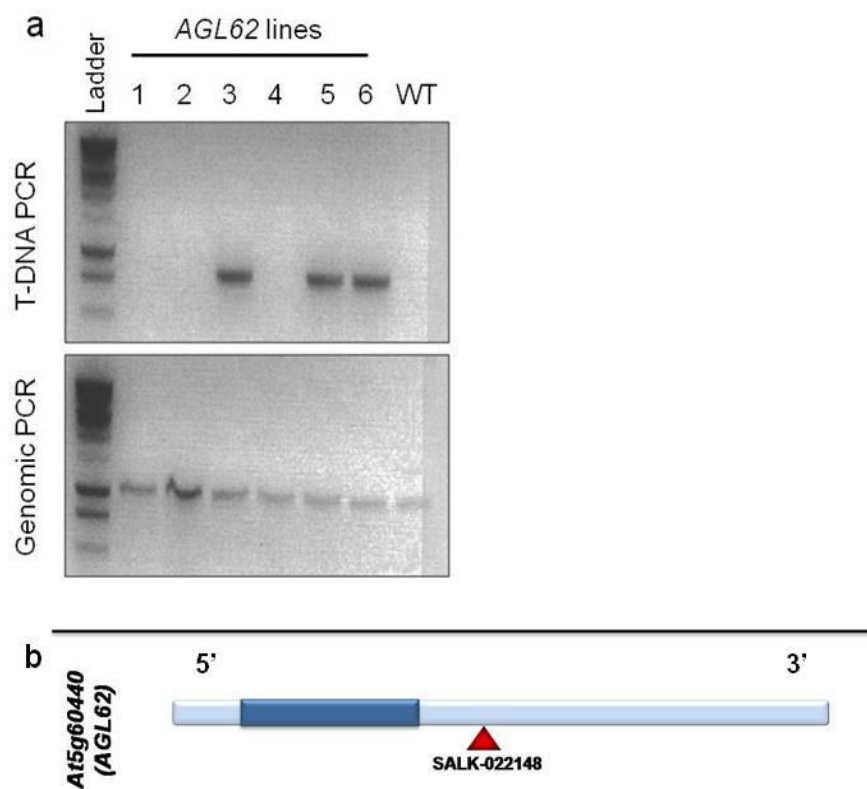


Figure 5.1 Analysis of the *AGL62* T-DNA insertion line, SALK-022148. (a) Genotyping: each lane represents a PCR product from DNA extracted from plants grown from the original seed supplied by the stock centre; heterozygous individuals had a band in both the T-DNA and genomic PCR; the wild types had a band only in the genomic PCR. The wild type (WT) was Col-0. DNA ladder is 1 kb. (b) Schematic showing the T-DNA insertion position in the SALK line. The light blue = exon; dark blue = introns; red triangle marks the position of the T-DNA insertion site.

5.2.1.2 Seed set of *AGL62/agl62* and wild type plants

The cause of the failure to recover *agl62/agl62* homozygotes from both the original seed lot and selfed heterozygotes was first investigated by analysing seed set in *AGL62/agl62* heterozygous plants. Seed in young siliques (10-12 day after pollination (DAP)) produced from *AGL62* heterozygous and wild type (Col-0) plants were examined under the dissecting microscope for any morphological differences to wild type seed and to assess seed set. *AGL62/agl62* siliques contained two visually very different classes of seed: seed of the first, more numerous class, were plump, green and indistinguishable from wild type, whilst seed of the second ‘defective’ class were extremely small, white, and shrivelled (Figure 5.2a-c). *AGL62/agl62* siliques contained 25% defective: 75% normal ($p < 0.001$) compared to wild type Col-0 that contained 100% normal seeds at the same stage of silique development (Figure 5.2d). Taken together with the genotyping results, this data strongly suggests that the failure to recover the *agl62/agl62* from selfed *AGL62/agl62* heterozygotes was caused by a failure of *agl62/agl62* seed to develop beyond an early stage. This indicated that *AGL62* plays a vital role in early seed development.

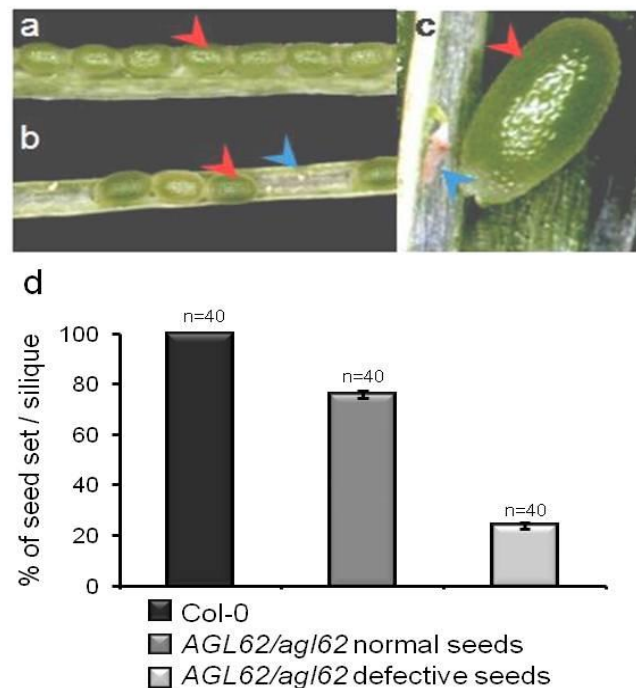


Figure 5.2 Seed phenotype analysis of *AGL62* heterozygous mutant and wild type plant. Images of seeds in green siliques (a) a Col-0 silique showing a normal seed phenotype, (b) An *AGL62/agl62* silique showing a mixture of seeds (‘normal’ and ‘defective’) (c) A magnified image of the normal and the defective seed. Red arrowhead = normal seed, blue arrowhead = defective seed. (d) Columns represent the percentage of seed set (normal or defective seeds) of *AGL62* heterozygous mutant and wild type (Col-0) plants, n = number of siliques examined, error bar = S.E.M.

5.2.1.3 *AGL62* segregation

In order to further investigate the cause of seed lethality in *AGL62/agl62* plants, reciprocal crosses between *AGL62/agl62* heterozygotes and wild type Col-0 (*AGL62/AGL62*) plants were made to determine the frequency of allele transmission. This would assess whether any parental differences existed in the transmission of the *agl62* allele. F1 seeds from both crosses were collected, sown and the resulting plants genotyped (Section 5.2.1.1). Both reciprocal F1 crosses segregated ~50% heterozygotes and ~50% wild type (Table 5.1). As expected no *agl62* homozygous plants were recovered (Table 5.1). This data shows that there was no parental bias in the transmission frequency of the *agl62* allele.

Table 5.1 Transmission of the *agl62* allele

Parental genotype		Progeny genotype		
Female	Male	<i>AGL62/AGL62</i>	<i>AGL62/agl62</i>	<i>agl62/agl62</i>
<i>AGL62/agl62</i>	<i>AGL62/AGL62</i>	52.5% (n=21)	47.5% (n=19)	0% (n=0)
<i>AGL62/AGL62</i>	<i>AGL62/agl62</i>	46.6% (n=20)	53.4% (n=23)	0% (n=0)

All seed present in a silique were collected; n= number of seed examined.

5.2.1.4 Loss of *agl62* function results in early seed lethality

In order to understand more about the seed defect phenotype associated with *AGL62* loss-of-function, the seed-clearing technique was applied to female gametophytes and to developing seeds at different times after pollination. A confocal microscopy analysis was also conducted on *AGL62/agl62* heterozygous developing seeds.

Since the timing of seed abortion in *agl62* mutant seed was early, developing seeds of manually selfed *AGL62* heterozygous mutant and wild type (Col-0) plants were examined at 0 to 4 DAP. Figure 5.3, shows that at 0 DAP (ovules before fertilization), ovules within *AGL62* heterozygous pistil had wild type appearance indicating no obvious defect in *AGL62/agl62* mutant female gametophyte. Similarly, at 1 DAP all seed appeared normal and were at very similar embryo development stages (1- 2 cells pro-embryo stage). In contrast, at 2 DAP the defective seed class became apparent in *AGL62/agl62* siliques; these seed were small compared to wild type, with embryo development typically at the quadrant embryo stage (4 cells) compared to normal seed within the same pod, or seed in wild type

pods, that were typically at the octant stage (8 cells). At 3 and 4 DAP the embryos of defective seeds were arrested at the octant embryo stage and the seeds were collapsed, whereas the wild type embryos were at globular and heart stage respectively.

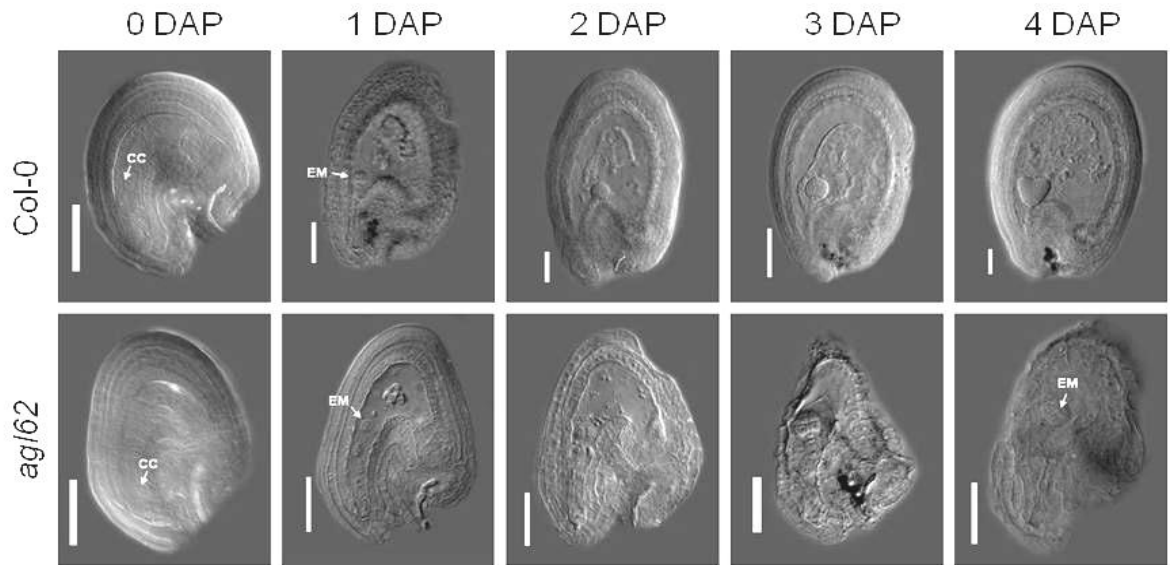


Figure 5.3 Seed phenotype analysis of wild type and *agl62* mutant seed by light microscopy. Developing seed of wild type (Col-0) and defective *agl62* seed at different times of embryo development. The female gametophyte (0 DAP) showed fused central cells (CC) in wild type and all seed examined in the *AGL62/agl62* siliques. 1 DAP seed showing an embryo (EM) at the 1-2 cell stage in both Col-0 and *AGL62/agl62*. At 2 DAP, the wild type seeds contained an octant-stage embryo, while a proportion of the *AGL62/agl62* seed were small and contained a 4-celled embryo. At 3 DAP the wild type embryo had reached the globular stage, whereas the embryo in *agl62* defective seed had arrested at the octant stage. The seed was also partially collapsed. At 4 DAP the wild type embryo was at a heart stage, while the embryo in *agl62* defective seed was obscured in a completely collapsed seed. DAP = Days after pollination, bar size = 50 μ m. Embryo development stages defined by Yadegari *et al.* (1994).

To investigate the seed phenotype further, particularly the behaviour of the endosperm, confocal microscopy was carried out on wild type and *AGL62/agl62* mutant siliques at different time after pollination. Figure 5.4 shows confocal images of wild type and *AGL62/agl62* seed at 2 DAP detailing the embryo and endosperm development stage. In wild type seed, the embryo was at the early globular stage and the endosperm was in the free nuclear form (Figure 5.4a-c). In contrast, in defective seed, the embryo was at octant stage and the endosperm was completely cellularised (Figure 5.4d-f). In wild type seed,

endosperm cellularisation seed begins at around 5 DAP when the embryo reaches heart stage (Sorensen *et al.*, 2001; Olsen, 2004). In general, the seed phenotype of the defective seed in the confocal images was similar to seed phenotype observed in cleared seeds and clearly showed that loss of *AGL62* function results in early seed lethality, that is associated a failure of the embryo to progress beyond the octant stage, and very precocious endosperm cellularisation. One hypothesis, given the expression profile of *AGL62*, is that the protein plays a crucial role in regulating the timing of endosperm cellularisation. Given the loss of function phenotype, its most likely role is to delay cellularisation until the seed has reached the appropriate developmental stage.

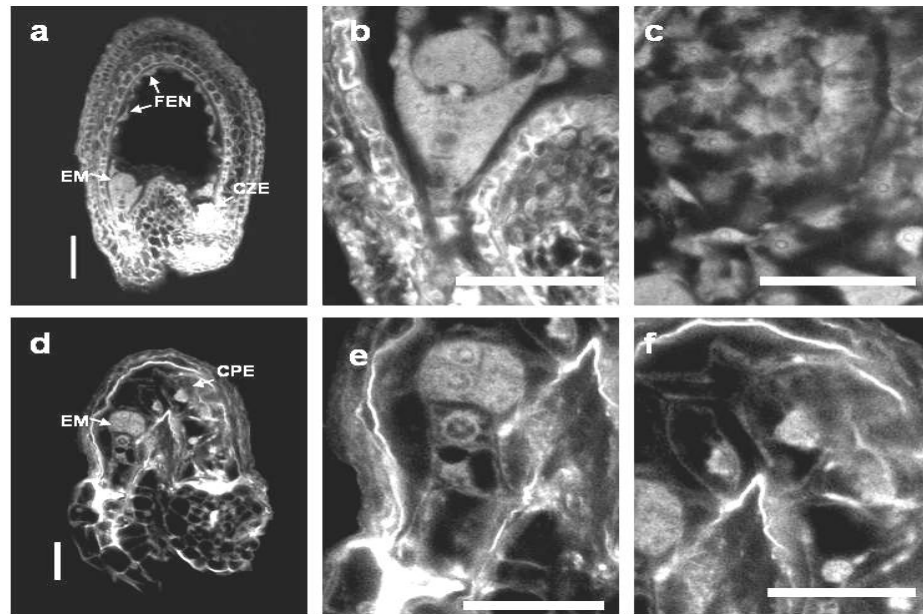


Figure 5.4 Confocal images of wild type and *agl62* defective seed at 2 DAP. Wild type Col-0 (a-c) and *agl62* mutant (d-f). (a) Wild type developing seed showed embryo (EM), chalazal endosperm (CZE) and free endosperm nuclei (FEN). (b) Wild type embryo at globular stage. (c) Wild type peripheral endosperm with free nuclei before cellularisation. (d) The *agl62* mutant defective seed showed an octant embryo (8-cells) and completely cellularised peripheral endosperm (CPE). (e) Octant embryo of *agl62* mutant seed. (f) Cellularised peripheral endosperm of *agl62* mutant seed. Bar size = 50 μ m.

5.2.2 Analysis of *AGL62* over-expression lines

The KI strategy was designed to produce over-expression of the target gene in the endosperm. Where the target gene is a suspected growth promoter the expectation is that this will result in endosperm over-proliferation leading to increased seed size. As explained earlier (Chapter 3; Section 3.1.1) over-expression of target genes in this study

was driven by the peripheral endosperm promoter (*pPER*) (Tiwari *et al.*, 2006). The *pPER* directs expression throughout the endosperm (peripheral, chalazal and micropylar regions), but not elsewhere in the plant (Tiwari *et al.*, 2006). The peripheral endosperm promoter was used to drive expression of *AGL62* coding region as this gene was positively associated with endosperm over-proliferation (Section 5.1.1) and therefore should be targeted to the whole endosperm.

5.2.2.1 Screening and selection of *AGL62* single insertion lines

The gene construct and *AGL62* over-expression lines were made by Dr. Tiwari (The Scott Laboratory, Bath) as described in section 3.2.3. The transformed *A. thaliana* (Col-0) plants were screened for the presence of the transgene (Figure 5.5). Lines with single insertions were identified from T2 plants using BASTA selection and subjected to T3 generation to select the homozygous lines (100% resistance to BASTA) (Chapter 2; Section 2.2.8).

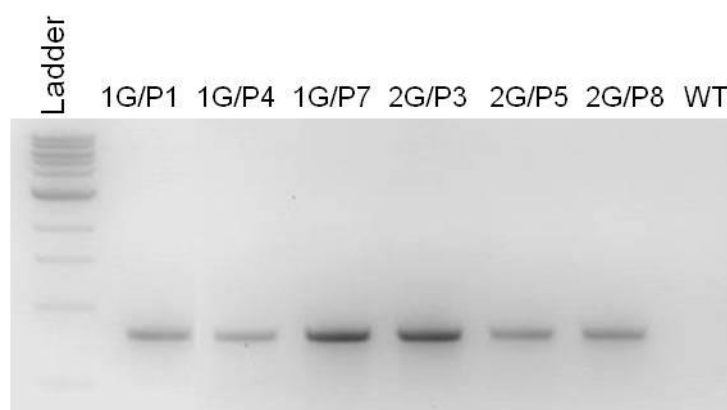


Figure 5.5 PCR products of *pPER::AGL62* lines. Each band represents a positive transformation. WT is wild type Col-0, ladder = 1kb. For DNA amplification, PER-LP and PER/at5g60440-RP primers were used (the primers and PCR conditions are detailed in Table 2.1 and Table 2.2, respectively).

5.2.2.2 Seed size analysis of *pPER::AGL62* plants

To determine whether over-expression of *AGL62* in the endosperm altered seed size, independent T3 homozygous *pPER::AGL62* lines were analysed for seed weight. No significant difference was found between the seed weight of five of *pPER::AGL62* transgenic lines (1G/P4/P7, $17.9 \pm 0.58 \mu\text{g}$; 1G/P7/P5, $20.4 \mu\text{g} \pm 1.1$; 1G/P1/P5, $19.9 \mu\text{g} \pm 0.21$; 2G/P8/P12, $21.3 \mu\text{g} \pm 0.87$ and 2G/P3/P5, $18.9 \mu\text{g} \pm 0.64$) and wild type ($19.6 \mu\text{g} \pm$

0.49) ($p = 0.082, 0.466, 0.660, 0.101$ and 0.454 , respectively). However, *pPER::AGL62-2G/P5/P12* plants did produce seed that was 10% larger ($21.4 \mu\text{g} \pm 0.30$) than wild type ($p=0.041$) (Figure 5.6a). Importantly, as shown in Figure 5.6b, only two transgenic lines were less fertile than wild type: *pPER::AGL62-1G/P4/P7* (43.6 ± 2.8) and *pPER::AGL62-2G/P5/P12* (42.6 ± 1.6) produced 21%; ($p=0.034$) and 23% ($p=0.019$) fewer seeds per silique than wild type (Col-0, 55.5 ± 2.8). The seed fertility of the other four *pPER::AGL62* transgenic lines: 1G/P7/P5 (60.3 ± 9.2), 1G/P1/P5 (58.7 ± 0.7), 2G/P8/P12 (59.3 ± 4.6) and 2G/P3/P5 (48.7 ± 2.9), was similar to wild type ($p = 0.528, 0.466, 0.476$ and 0.174 , respectively). The work stopped at this point because of lack of time.

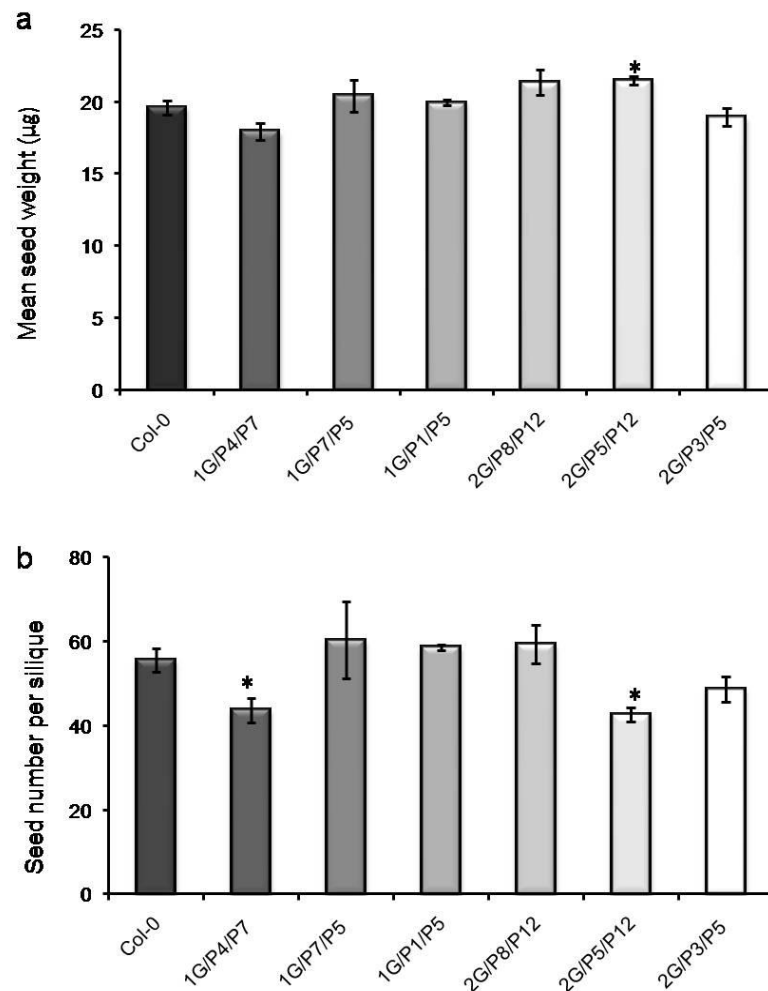


Figure 5.6 Seed weight analysis of six independent homozygous single insertion *pPER::AGL62* lines. (a) Each bar represents the mean seed weight ($n \geq 250$ seeds). (b) Each bar represents the mean number of seed set per silique ($n = 5$ siliques/3 plants). For comparisons, values that differ at the 0.05 significance level from wild type are labelled with a *. Error bar = S.E.M.

5.3 Discussion

During the course of the work reported in this chapter the Drews laboratory (USA) published the findings of similar work on the SALK-022148 and SALK-13770 insertion lines (Kang *et al.*, 2008). In agreement with the findings above, this group reported that the AGL62 protein regulates the timing of endosperm cellularisation during early seed development in *A. thaliana*.

5.3.1 AGL62 mutation affects seed development

There was a failure to recover any homozygous *agl62* individuals from the AGL62 T-DNA insertion line (SALK-022148) or from a large number of plants derived from self pollinated heterozygotes. Analysis of young siliques produced by AGL62 heterozygous plants revealed that a proportion of the seed suffered early seed abortion, strongly suggesting that homozygosity of the *agl62* allele is lethal. This idea was supported by genotyping a larger number of F2 progeny from selfed AGL62/*agl62* plants which again resulted in only the AGL62/AGL62 and AGL62/*agl62* genotypes in a 1:2 ratio.

To determine whether *agl62* had any effect on gametic transmission, reciprocal crosses made between AGL62 heterozygous and wild type plants were analysed. The F2 progeny for both crosses contained AGL62/AGL62 and AGL62/*agl62* genotypes only, in a 1:1 ratio, indicating that the female gametophyte was not affected by the AGL62 mutation and that the *agl62* allele could be transmitted by both parents. These findings are in agreement with those of Kang *et al.* (2008). In conclusion, our findings reported here and those of the Drews laboratory suggest that AGL62 loss-of-function mutation causes recessive seed lethality.

5.3.2 AGL62 controls early endosperm cellularisation and possibly will promote proliferation

In order to understand how AGL62 loss of function causes seed lethality a thorough microscopy analysis was carried out on early seed development. There were no observable defects of the female gametophyte. In contrast, a proportion of developing seeds in self pollinated AGL62/*agl62* siliques displayed very obvious developmental differences to wild type seed from 2 DAP onward. The most striking features of these seeds were embryo arrest at the 8-cell stage and very early cellularisation of the endosperm (Figures 5.3 and 5.4).

Kang *et al.* (2008) also report the results of confocal microscopy as well as counts of endosperm nuclei number in SALK-13770 which showed similar phenotype to SALK-022148. *agl62* mutant seeds showed three types of defects: precocious endosperm cellularisation, reduced number of endosperm nuclei and abnormal embryo development. However, there was one difference between our data and that of Kang *et al.* (2008) relating to the timing of embryo arrest in mutant seeds: we found this occurred at the octant embryo stage (8 cells) whilst they reported arrest at the quadrant stage.

Kang *et al.* (2008) used a *pAGL62-GFP* reporter gene to demonstrate that *AGL62* expression within the seed is restricted to the endosperm. This supports our suggestion (5.2.1.4) that the arrest of embryo development is an indirect consequence of highly precocious endosperm cellularisation, which would appear to be the primary effect of *agl62* loss of function. Hypothetically, early endosperm cellularisation could result from a reduction in nuclear proliferation or *vice versa*. Since several cycles of cell division occur after the initiation of cellularisation in the wild type seed (Brown *et al.*, 1999), cellularisation does not prevent mitosis. Consequently, the small size of the *agl62* endosperm suggests that *AGL62* might play a role in endosperm mitosis. Under this scenario, the few early rounds of mitosis within *agl62* seeds could be explained by the presence of *AGL62* protein synthesized prior to meiosis from the sporophytic genome and deposited in the embryo sac.

Precocious endosperm cellularisation does not necessarily result in a seed lethal phenotype. The endosperm of the *A. thaliana* mutants *iku1*, *iku2* and *mini3* cellularise between 1-3 DAP, but produce viable seed (Garcia *et al.*, 2003; Luo *et al.*, 2005). The *iku* seeds produced a normally developed embryo (bent cotyledon) but the endosperm cellularised early which was suggested to be responsible for the small seed phenotype (Garcia *et al.*, 2003). This suggests that *agl62* is much more severe than previously described mutants that affect the timing of endosperm cellularisation, and that there is a lower limit to endosperm size, below which the seed does not survive.

As described in (Chapter 1; Section 1.5), a striking feature of the parent-of-origin effect associated with interploidy crosses is altered timing of endosperm cellularisation. In *A. thaliana* paternal excess (2x X 4x and 2x X 6x) crosses result in delayed or absent endosperm cellularisation, whilst endosperm cellularisation in maternal excess (4x X 2x and 6x X 2x) crosses is precocious (Scott *et al.*, 1998; Xiao *et al.*, 2006). Since genomic imprinting underpins parent of origin effects, genes controlling the timing of endosperm

cellularisation, such as *AGL62*, are likely targets for imprinting. However, our data on *AGL62* segregation patterns (Sections 5.2.1.2 and 5.2.1.3) together with *pAGL62-GFP* expression analysis (Kang *et al.*, 2008) show that *AGL62* is expressed from both parental alleles in endosperm, and is therefore not subject to genomic imprinting. Nevertheless, *AGL62* could be regulated directly or indirectly by imprinted genes, and therefore play a role in the parent of origin effect.

5.3.3 Over-expression of *AGL62* in endosperm has no effect on final seed size

The second strategy to test the role of *AGL62* in endosperm growth and development involved its over-expression in the early endosperm. Our transcriptional profiling experiments identified *AGL62* as a potential growth promoter, which was consistent with the outcome of the KO analysis that showed that loss of function resulted in reduced seed size. Since *AGL62* appears to delay endosperm cellularisation, and delayed endosperm cellularisation is a feature of paternal excess (see above), we reasoned that *AGL62* over-expression in the endosperm might mimic paternal excess by delaying cellularisation, and thereby increase seed size. However, introduction of a *pPER::AGL62* construct into *A. thaliana* failed to provide a consistent increase in seed weight (5.2.2.2). From six individual independent homozygous T3 *pPER::AGL62* plants, only one transformed plant (*pPER::AGL62*-2G/P5/P12) produced 10% heavier seeds, but with relatively low seed number per pod; lines with normal levels of seed fertility had near wild type seed size (Figure 5.6).

The failure to increase seed size by transforming plants with *pPER::AGL62* may be explained in a number of ways. First, we did not determine whether the construct increased the level of *AGL62* at the critical developmental stage (4-5DAP) and therefore cannot say with certainty that over-expression was achieved. Second, since *AGL62* appears to interact with other proteins, namely PHE1, PHE2, AGL40, AGL45 and AGL28 (de Folter *et al.*, 2005; Section 5.1.1), which may be required to operate within multi-protein complex to block cellularisation, over-expression of only one component (*AGL62*) would therefore be insufficient to delay cellularisation.

5.3.4 Future work

As discussed above, the introduction of the *pPER::AGL62* construct alone into plants did not affect seed weight. Potential improvements to the strategy used in the present work include: 1) confirming that the construct drives increased *AGL62* transcript levels beyond

the normal time of endosperm cellularisation; 2) introducing the *pPER::AGL62* into lines over-expressing one or more of its potential interaction partners. The Scott laboratory has produced a number of suitable lines e.g. *pPER::PHE1*, *pPER::PHE2*, *pPER::AGL40*, *pPER::AGL45* and *pPER::AGL28* constructs (Chapter 3; Hughes, 2009) which could be stacked with *pPER::AGL62* by hybridisation.

Chapter 6: Verification of genes involved in endosperm growth and development using a knock-Out (KO) strategy

6.1 Introduction

In Chapter 1 we discussed the correlation between seed and endosperm growth in crop plants and the importance of increasing seed size by manipulating endosperm size in the model plant *Arabidopsis thaliana* (*A. thaliana*). In previous chapters we verified the role of the *phe1* homozygous mutant in controlling seed size and the *agl62* homozygous mutant in causing seed lethality. The study in this chapter focuses on verifying the role of other candidate genes, namely: *PHERES 2* (*PHE2*), *AGL40*, *AGL45* and *AGL28*, which are grouped under the MADS box category and are associated with endosperm proliferation. We were successful in over-expressing some of the MADS box candidate genes in our laboratory, namely: *PHERES 1* (*PHE1*) and *AGL28* (Hughes, 2009), *PHE2*, *AGL40*, *AGL45* and *AGL62* (Chapters 3 and 5), and their effect on seed size was demonstrated and subsequently discussed. In this chapter, the study focuses on investigating the role of these genes in influencing seed size by means of a knock-out (KO) strategy where the target genes were inactivated by a means of T-DNA insertion.

6.1.1 Genes associated with endosperm over-proliferation

The Scott laboratory had previously used a microarray approach to analyse the expression of seed genes in different interploidy crosses relative to the balanced 2x X 2x cross (Tiwari *et al.*, 2010). Genes displaying large changes in relative expression, either increasing or decreasing, were listed as candidate genes potentially responsible for the observed differences in endosperm size and the differentiation in the seeds generated in the various crosses (see Chapter 1). This list contained a large number of candidate genes that could have been verified by further experiment. However, given the limited resources, a coherent subset of genes was sought for verification. We had previously confirmed that the MADS box genes *PHE1* (*AGL37*) and *AGL62* (Chapters 4 and 5) played important roles in endosperm development. Since the long list of candidate genes that were highly expressed in over-proliferated endosperms and positively associated with endosperm development included the MADS-box genes *PHE2* (*AGL38*), *AGL45*, *AGL40* and *AGL28*, we chose these genes for further investigation (see Chapter 3). The *Arabidopsis* genome has 107 genes which possess a MADS-box domain and represent a large family of transcription

factors, of which 84% have unknown functions ((Parenicova *et al.*, 2003). The MADS family consists of two groups: type I MADS box proteins (~61), which can be subdivided into the M α , M β , and M γ types; and type II MADS box proteins which contain the MIKC domain (MADS-box-intervening-keratin-like C-terminal region) and the M δ types (Alvarez-Buylla *et al.*, 2000a; Parenicova *et al.*, 2003). Several genes belonging to the MADS box family have been reported to have allocated roles in different aspects of plant development, such as plant reproduction, flowering time, floral organogenesis, floral meristem identity, fruit formation, seed pigmentation and endothelium development (Ng and Yanofsky, 2001; Nesi *et al.*, 2002; Becker and Theissen, 2003).

The candidate genes belong to a type I MADS box and their coding region contains a highly conserved putative AGAMOUS-LIKE MADS domain (AGL) which binds DNA, and they were considered as transcription factors (Alvarez-Buylla *et al.*, 2000a; Alvarez-Buylla *et al.*, 2000b; Parenicova *et al.*, 2003). Also, Nam *et al.* (2003) suggested in their study that the type I genes may be more variable among different angiosperm species than the type II genes because of their faster birth-and-death evolution than type II genes. In addition, type I genes are generally less conserved than type II genes. Few AGLs that belong to this group (type 1) have been reported with known functions in plant development to date. The first gene to be characterized, *PHE1* (AGL37), an M γ type gene, was found to be regulated by the Polycomb group gene *MEDEA* (*MEA*) and expressed in embryos and endosperms shortly after fertilization (Kohler *et al.*, 2003b), and the loss-of-function mutant (*phe1.1*) showed a small seed phenotype compared to the wild type (Chapter 4). More evidence for the functionality of M γ -type genes was provided by the *AGL80* loss-of-function mutant (*agl80*) (Portereiko *et al.*, 2006). The *agl80* mutant megagametophytes show a defect in the maturation of the central cell and fail to develop endosperms after fertilization. In conformity with the mutant phenotype, *AGL80* is expressed in the central cell and in the endosperm before cellularisation. The M α -type gene *AGL62* was found to be expressed during the syncytial phase of endosperm development (Kang *et al.*, 2008). The *AGL62* loss-of-function mutant (*agl62*) endosperm undergoes precocious cellularisation, indicating that *AGL62* plays an important role in endosperm development (see also Chapter 5). Finally, the M α -type gene *AGL23* was found to be involved in both female gametophyte and embryo development (Colombo *et al.*, 2008). The functional studies of *PHE1*, *AGL80*, *AGL62* and *AGL23* imply a role for type I MADS box genes in female gametophyte development and/or early seed development. Therefore, it is important to study the functions of the other genes grouped within this family either as

single genes or within a complex, and because MADS domain proteins classically function as homo- or heterodimers with other MADS-domain proteins and often in higher-order complexes, so they may have similar crucial effects on plant development. In addition, the proteins of the candidate genes in this study have been reported to interact with each other either directly or indirectly; the protein-protein interaction map shown in Figure 6.1, derived from de Folter *et al.* (2005), illustrates the proposed interactions between the candidate proteins.

The objective of this study was to investigate the role of the candidate genes, namely: *PHE2*, *AGL40*, *AGL45* and *AGL28*, in endosperm growth and development and their effects in altering the final seed size in the model plant *A. thaliana* as single genes or in complexes by applying a KO strategy to the target genes to allow gene inactivation and loss of function.

As explained earlier, the experimental strategy proposed to investigate the functions of candidate genes was designed to cause inactivation of the target gene. Therefore, hypothetically, the KO of growth promoter genes was expected to produce smaller seeds, as discovered in *phe1.1* mutant phenotype, while the KO of growth inhibitor genes was expected to give rise to larger seeds. The exact outcome was expected to depend on whether the gene was endosperm specific, seed specific or expressed generally.

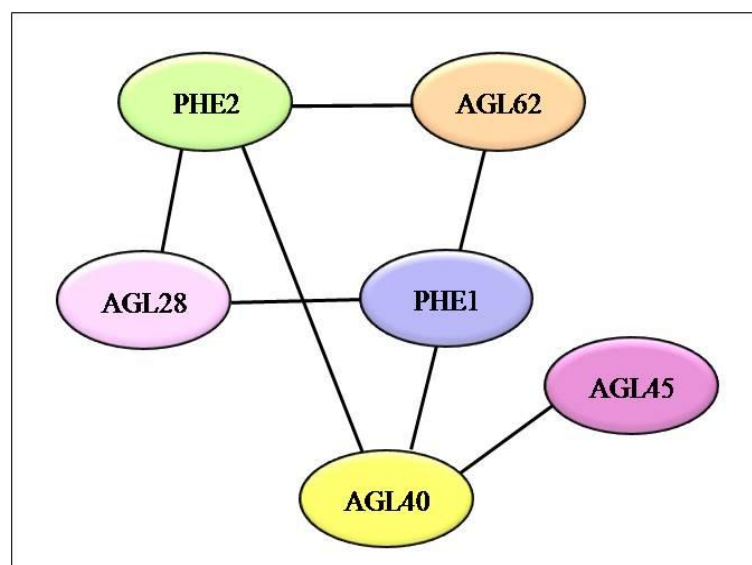


Figure 6.1 Protein-protein interaction map of the candidate MADS box proteins in a yeast two-hybrid system. Each protein is highlighted by a different colour. The lines indicate the interactions between the genes. PHE2 interacts with AGL40, AGL28 and

AGL62; PHE1 interacts with AGL40, AGL62 and AGL28; AGL40 interacts with AGL45; AGL62 interacts with PHE2 and PHE1.

6.2 Results

6.2.1 Genotyping T-DNA insertion lines

In order to verify the genotype of the insertion lines before engaging in further experiments, plants grown from seeds obtained from the stock centre were genotyped using a PCR-based procedure that detects the presence of T-DNA in the target gene. The PCR products were verified by gel electrophoresis and the bands were visualized and compared with the DNA ladder to determine the fragment sizes (the primers and PCR conditions are detailed in Table 2.1 and Table 2.2, respectively). The results showed that all of the insertion lines examined contained a T-DNA in the expected genes and this was confirmed by observing homozygous and heterozygous genotypes illustrated in (Figure 6.2); only homozygous lines were used further in this study.

6.2.2 Sequencing T-DNA homozygous insertion lines

In order to confirm the position of the T-DNA within the target gene the PCR products were verified by DNA sequencing. The sequence was viewed and analysed by Sequencher 4.9 Demo and compared with the original data in the SIGnAL website (<http://signal.salk.edu/>). All of the insertion lines T-DNAs were confirmed as being in the positions detailed on the SIGnAL website (Figure 6.3).

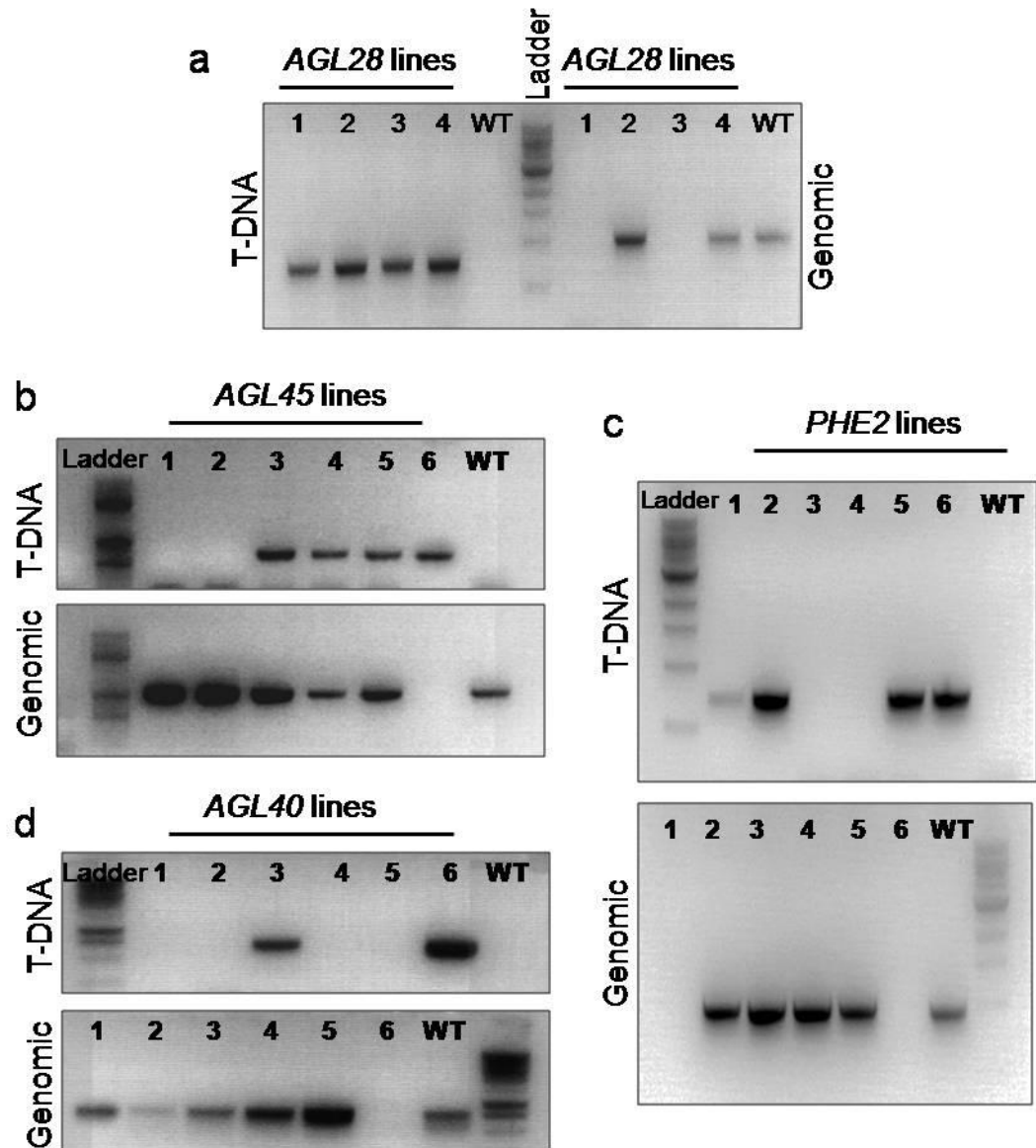


Figure 6.2 Genotyping of the T-DNA insertion lines. Each lane represents a PCR product of DNA extracted from plants grown from different seeds obtained from the stock centre. Homozygous lines only showed bands in the T-DNA PCR; heterozygous lines showed bands in both the T-DNA and genomic PCR; the wild type only showed bands in genomic PCR. (a) *AGL28* T-DNA insertion lines: lanes 1 and 3 were homozygous, lanes 2 and 4 were heterozygous, WT was Ws-4. (b) *AGL45* T-DNA insertion lines: lanes 1 and 2 were wild type, lane 6 was homozygous and lanes 3, 4 and 5 were heterozygous; WT was Col-0. (c) *PHE2* T-DNA insertion lines: lanes 1 and 6 were homozygous, lanes 2 and 5 were heterozygous, lanes 3 and 4 were wild types, and WT was Col-0. (d) *AGL40* T-DNA insertion lines: lanes 1, 2, 4 and 5 were wild types; lane 3 was heterozygous and lane 6 was homozygous; WT was Col-0. The DNA ladder was 1 kb.

6.2.3 Embryo sac area measurements of the homozygous T-DNA insertion mutants

This study focused on endosperm development and its correlation with final seed size. In order to test the roles of the target genes in endosperm development, it was important to measure endosperm size in the insertion lines. However, since counting the number of endosperm nuclei is very time consuming, we measured the embryo sac area as this is a good proxy for endosperm size (Scott *et al.*, 1998). The embryo sac area refers to the whole area surrounded by the seed coat and containing the embryo, the peripheral endosperm and the chalazal endosperm cysts. Developing seeds were collected from homozygous mutants and wild types and the embryo sac areas were measured and analysed to determine endosperm size.

6.2.3.1 Seed size analyses of the single homozygous T-DNA insertion plants

6.2.3.1.1 Embryo sac area measurement

In order to test the roles of *PHE2*, *AGL40*, *AGL45* and *AGL28* in endosperm development, developing seeds of the single homozygous T-DNA insertion plants; *phe2.1* (Col-0 background), *agl40.1* (Col-0 background), *agl45.1* (Col-0 background) and *agl28.1* (Ws-4 background); and their wild types, were collected at different times after pollination (DAP). After collection the seeds were cleared and then photographed; the embryo sac area of the photographed seeds was measured and the mean embryo sac area was analysed statistically by either the Student's *t*-test or the Mann-Whitney *U*-test at a 95% confidence level.

Figure 6.4 shows no difference in the seed morphology of *phe2/phe2* compared to the wild type (Col-0) at 4, 5, 6 and 7 DAP. Although, the embryo sac area of the *phe2/phe2* seed at 4 DAP was significantly larger (3.5%) than Col-0 at the same time ($p = 0.011$), the area measured at 6 DAP was significantly smaller (13%) compared to the wild type ($p < 0.001$). Conversely, the embryo sac areas at 5 and 7 DAP were not found to be significantly different compared to the wild type ($p = 0.155$ and 0.891 , respectively) (Figure 6.5). Therefore, there was no consistent difference between the wild type and *phe2/phe2*, suggesting that *phe2.1* KO does not affect endosperm size.

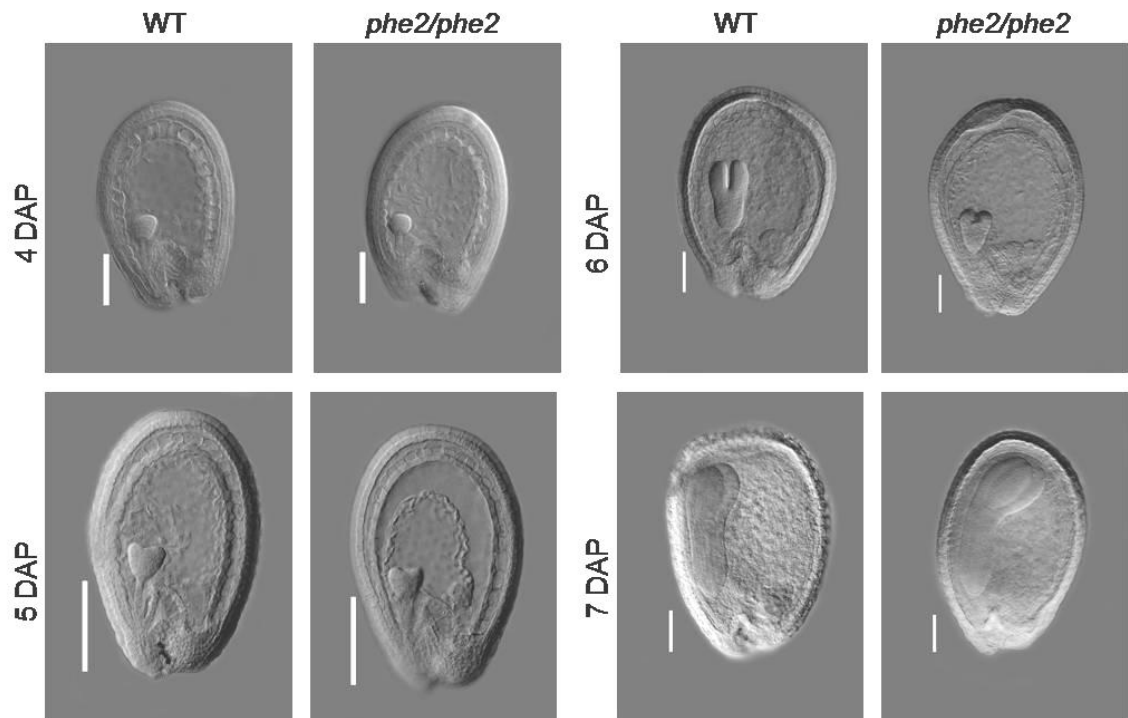


Figure 6.4 Seed development in wild type and *phe2/phe2* plants. Cleared, developing seeds were imaged using differential interference contrast (DIC) optics at 4, 5, 6 and 7 DAP. The wild type (WT) is Col-0. Bar size = 100 μm .

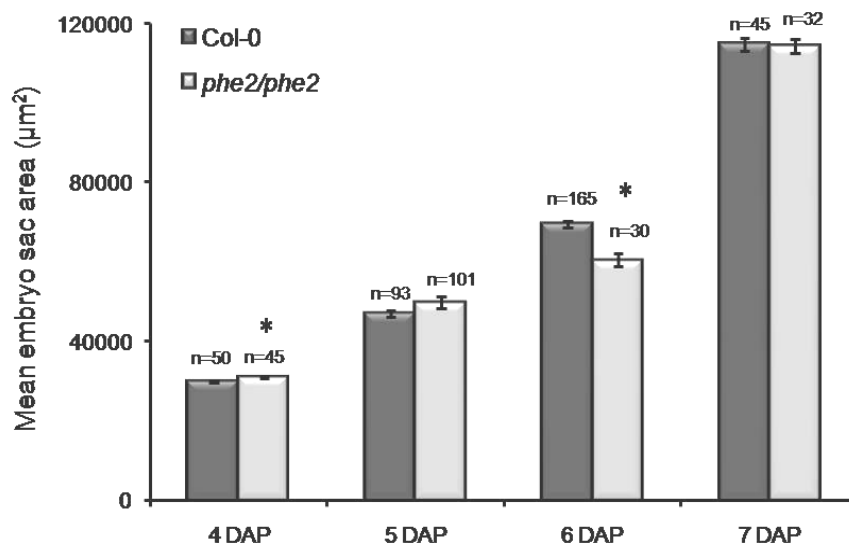


Figure 6.5 Embryo sac area measurements of wild type and *phe2/phe2* plants. The columns represent the mean areas of the wild type (Col-0) and the homozygous mutant (*phe2.1*) at 4, 5, 6 and 7 DAP. For comparison, the values which differed at the 0.05 significance level between the wild type and the homozygous mutant are labelled with *. Error bar = S.E.M.; n = number of seeds examined.

The developing seeds of *agl40/agl40* were subjected to the same analysis as described above (illustrated in Figure 6.6). Although a difference was observed in the seeds regarding the embryo developmental stage morphology of *agl40/agl40* at 3 and 5 DAP compared to the wild type, at 7 DAP both the mutant and wild type seeds had reached the torpedo embryo stage. The embryo sac area of the *agl40.1* homozygous mutant was significantly larger than the wild type at 3 and 5 DAP (Figure 6.7): 22.3% ($p < 0.001$) and 12.6% ($p < 0.001$), respectively. In contrast, no significant differences were observed in the embryo sac area sizes at 4, 6 and 7 DAP between the homozygous mutant and the wild type ($p = 0.427$, 0.213 and 0.800 , respectively). Since there was no consistent difference, this evidence suggests that the KO of *agl40.1* has no effect on seed or endosperm development, as found for the *phe2.1* KO mutant.

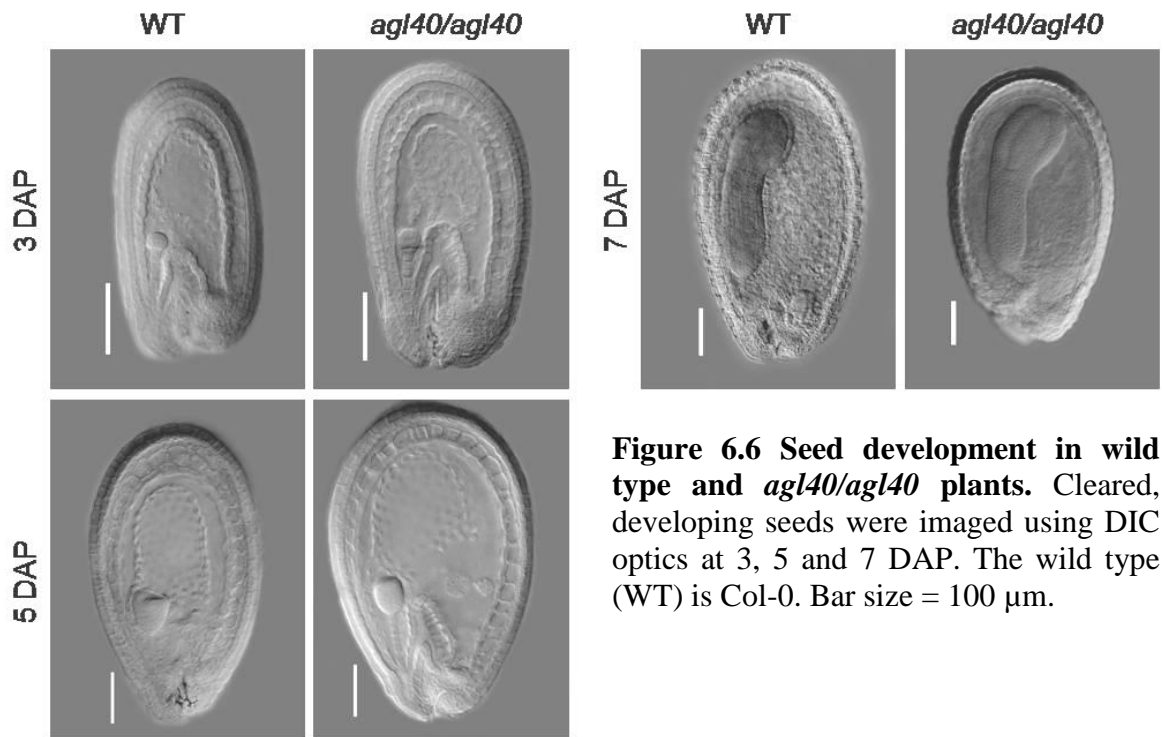


Figure 6.6 Seed development in wild type and *agl40/agl40* plants. Cleared, developing seeds were imaged using DIC optics at 3, 5 and 7 DAP. The wild type (WT) is Col-0. Bar size = 100 μm .

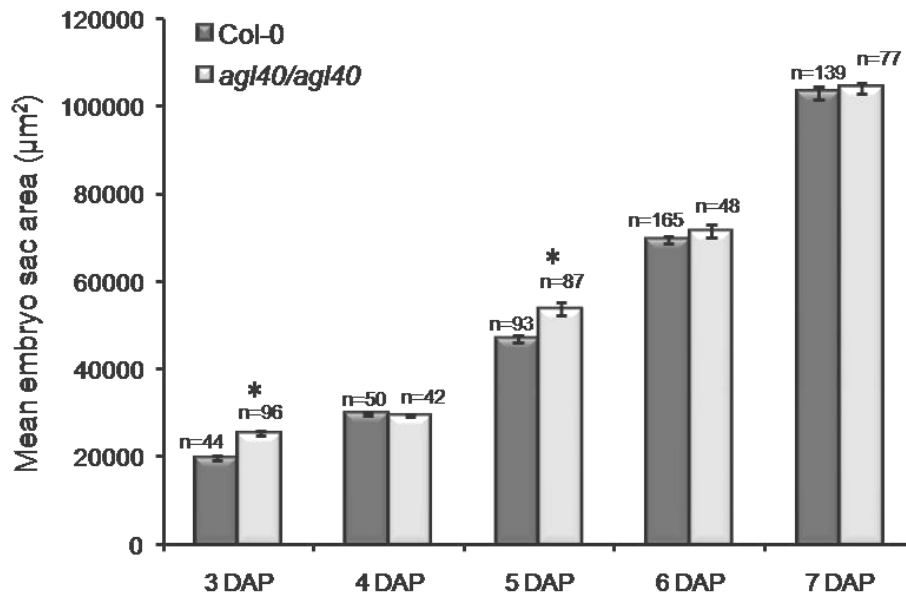


Figure 6.7 Embryo sac area measurements of wild type and *agl40/agl40* plants. The columns represent the mean areas of the wild type (Col-0) and the homozygous mutant (*agl40.1*) at 3, 4, 5, 6 and 7 DAP. For comparison, the values which differed at the 0.05 significance level between the wild type and the homozygous mutant are labelled with *. Error bar = S.E.M.; n = number of seeds examined.

The third candidate T-DNA insertion mutant was *agl45/agl45*. As with the previous mutant lines, the developing, cleared seeds were imaged at different time intervals and displayed no differences in embryo stage when compared to the wild type seeds (Figure 6.8). The embryo sac area of the *agl45/agl45* mutant plant was not significantly different compared to the wild type (Col-0) plant at any of the time points measured (4, 6 and 7 DAP) ($p = 0.115, 0.879$ and 0.651 , respectively) (Figure 6.9).

Finally, the fourth candidate mutant was *agl28/agl28*. The developing seeds of these single mutant plants were cleared and photographed at different time points after pollination along with seeds of the wild type, which in this case was Ws-4. There were no observable differences (Figure 6.10). In addition, there were no significant differences in the embryo sac areas at 4 or 5 DAP ($p = 0.890$ and 0.301 , respectively) (Figure 6.11).

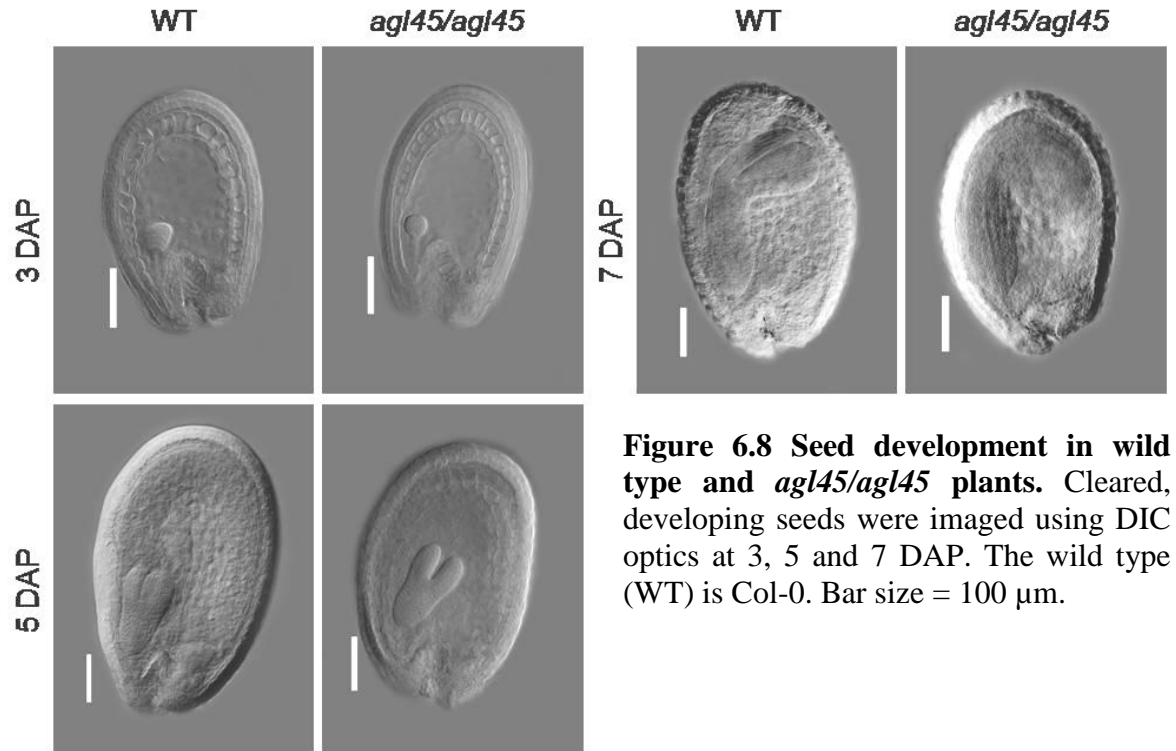


Figure 6.8 Seed development in wild type and *agl45/agl45* plants. Cleared, developing seeds were imaged using DIC optics at 3, 5 and 7 DAP. The wild type (WT) is Col-0. Bar size = 100 μm .

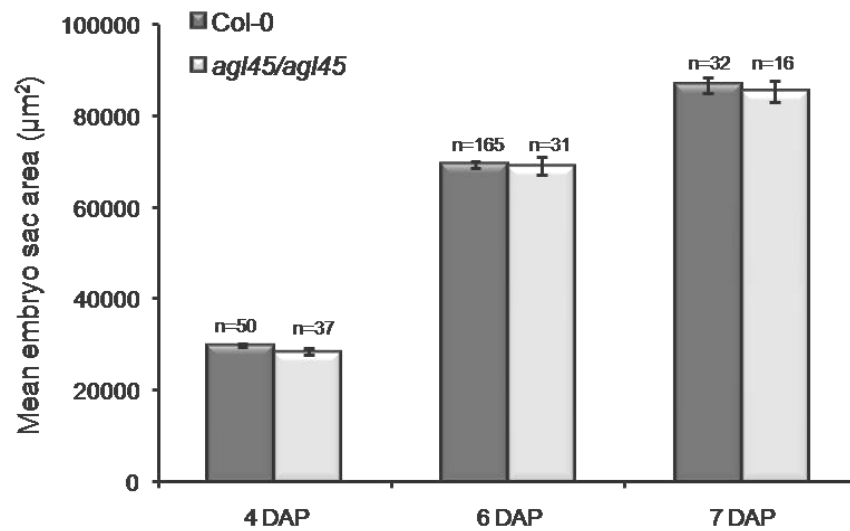


Figure 6.9 Embryo sac area measurements of wild type and *agl45/agl45* plants. The columns represent the mean areas of the wild type (Col-0) and the homozygous mutant (*agl45.1*) at 4, 6 and 7 DAP. Statistical analyses were measured at the 0.05 significance different level. Error bar = S.E.M.; n = number of seeds examined.

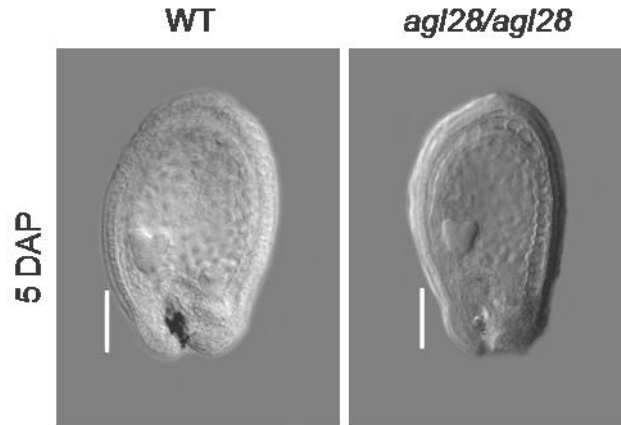


Figure 6.10 Cleared, developing seeds from wild type and *agl28/agl28* plants at 5 DAP. Cleared, developing seeds were imaged using differential interference contrast (DIC) optics. The wild type (WT) is Ws-4. Bar size = 100 μ m.

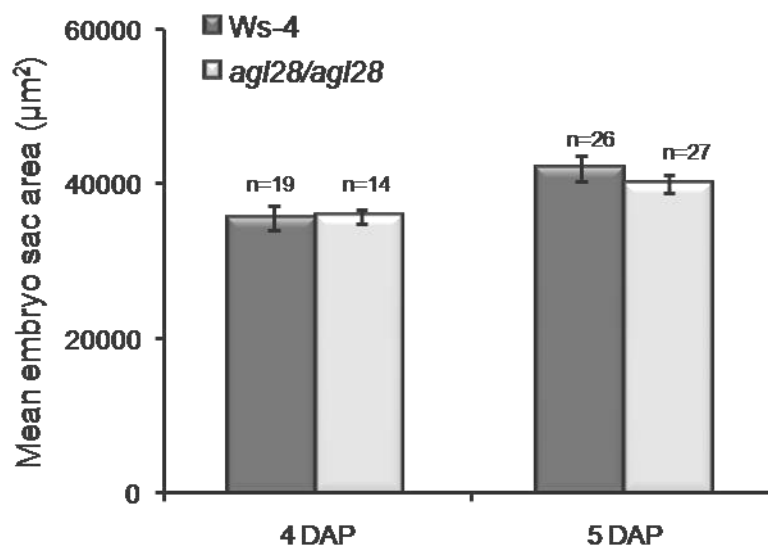


Figure 6.11 Embryo sac area measurements of wild type and *agl28/agl28* plants. The columns represent the mean areas of the wild type (Ws-4) and the homozygous mutant (*agl28.1*) at 4 and 5 DAP. Statistical analyses were measured at the 0.05 significance different level. Error bar = S.E.M.; n = number of seeds examined.

6.2.3.1.2 Individual seed weight analyses

The seed size in *A. thaliana* could become changed by altering endosperm growth after fertilization. The results obtained from the embryo sac area measurements indicate that the single homozygous KO mutants have a similar embryo sac area size to the wild type, as detailed above. Does the similarity in embryo sac area size translate into a similar final seed size? To answer this question, seed weight analyses were carried out to determine the relationship between embryo sac area and final seed weight. Mean seed weights were measured for *phe2/phe2*, *agl40/agl40*, *agl45/agl45* and *agl28/agl28* homozygous T-DNA insertion plants, and then the mean weights of individual seeds were calculated (Table 6.1). The measurements of seed weights obtained from *agl40/agl40*, *agl45/agl45* and *agl28/agl28* homozygous mutants compared to those of the wild type confirmed that there were no significant differences in the seed weight. In contrast, although there was no consistent difference in the embryo sac area size of the *phe2.1* homozygous mutant when compared to the wild type, the mean individual seed weight of the *phe2/phe2* mutant plants was significantly different, with *phe2/phe2* producing seeds that were 6% heavier than the wild type. One explanation for this is that the embryo sac area was only a measurement of endosperm size, whereas the final seed weight included contributions from the embryo, the endosperm and the seed coat (integument layer).

Table 6.1 Seed weight of the single homozygous T-DNA insertion lines and their wild types.

Genotype	Mean seed weight (μg) \pm S.E.M ^a	Number of seeds	Average seed number/pod	p-value
Col-0	19.1 \pm 0.37	2189	52	
<i>phe2/phe2</i> (Col-0)	20.4 \pm 0.28	2297	54	0.009 ^{b*}
<i>agl40/agl40</i> (Col-0)	19.7 \pm 0.17	2429	60	0.165 ^b
<i>agl45/agl45</i> (Col-0)	19.0 \pm 0.89	438	63	0.908 ^b
Ws-4	19.2 \pm 0.56	280	40	
<i>agl28/agl28</i> (WS-4)	19.7 \pm 0.21	406	58	0.433 ^c

Genetic background of the mutant plants shown in brackets

^a S.E.M seed weight of $n \geq 7$ siliques

^b Student's *t*-test with Col-0 wild type, at the 0.05 significant difference level.

^c Student's *t*-test with Ws-4 wild type, at the 0.05 significant difference level.

*Significantly different compared to wild type.

6.2.4 Analysis of double mutants

Individual knock-outs of *PHE2*, *AGL40*, *AGL45* and *AGL28* did not affect endosperm proliferation. One possible explanation for this lack of phenotypes is that functional redundancy exists for this set of genes. Functional redundancy is raised when the knock-out of a single gene gives no mutant phenotype, and is particularly likely where related duplicate genes exist (Bouche and Bouchez, 2001; Moore *et al.*, 2005). A good example of functional redundancy exists within the MADS box family *SEPALLATA1-3* (*SEP1-3*). These genes are involved in the patterning of the floral meristem (Pelaz *et al.*, 2000). Single *sep* mutants do not produce an observable phenotype, but flowers of the triple mutant, *sep1/sep2/sep3*, are entirely composed of sepals. A double mutant strategy was developed to test whether or not functional redundancy exists in the *PHE2*, *AGL40*, *AGL45* and *AGL28* set. The double homozygous mutant strategy was designed in two complexes as explained in Chapter 2, Section 2.2.7: one complex was centred around *phe2/phe2* and the other complex was centred around *phe1/phe1*. All possible combinations of these two complexes are illustrated in Figure (6.12). Crosses were made between single mutants to produce F1 hybrid seeds that were heterozygous for both genes; these were then segregated into F2 plants. The F2 plants were generated and genotyped to confirm the genotypes of the double complexes. All of the six double homozygous mutants (*phe2::agl40*, *phe2::agl45*, *phe2::agl28*, *phe1::agl40*, *phe1::agl45* and *phe1::agl28*) were similar to the wild type vegetative phenotypes, so we decided to choose one double mutant from each complex: *phe2/phe2::agl40/agl40* and *phe1/phe1::agl40/agl40* were selected for further analysis.

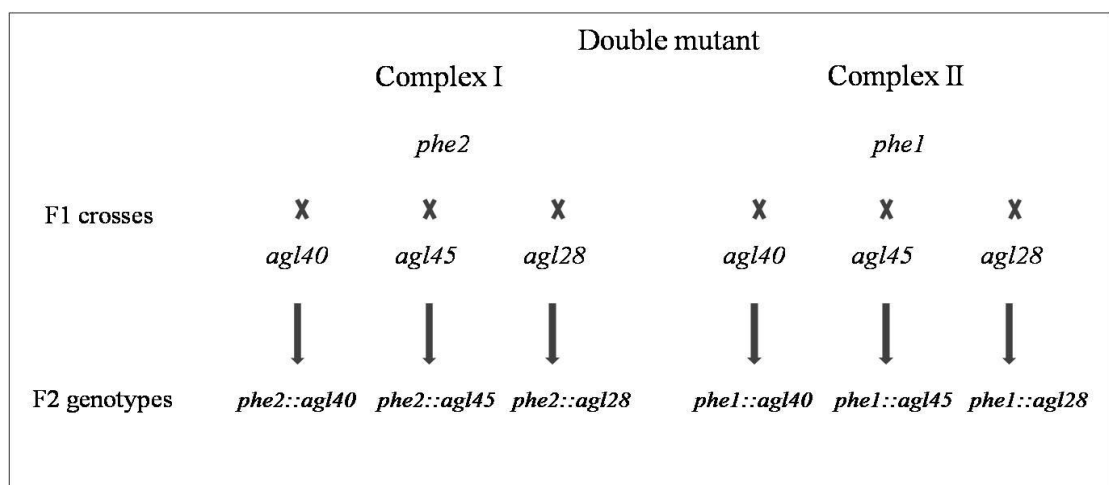


Figure 6.12 Diagram of the hypothetical double mutant combinations centred around *phe2* and *phe1* homozygous mutants.

6.2.4.1 Analyses of the *phe2/phe2::agl40/agl40* plants

In order to test the effect of the *phe2/phe2::agl40/agl40* double mutant on endosperm proliferation, crosses were made between *phe2/phe2* and *agl40/agl40* single mutant plants to produce F1 hybrid seeds that were heterozygous for both genes (*PHE2/phe2::AGL40/agl40*). Following this, F2 plants were generated and genotyped. The *phe2/phe2::agl40/agl40* double homozygotes were identified (Figure 6.13) and used for further analysis.

The embryo sac area was measured from cleared seeds at 5 and 7 DAP (Figure 6.14). At 5 DAP, the *phe2/phe2::agl40/agl40* seed was 21.7% larger than the wild type seed ($p < 0.001$). However, at 7 DAP there was no significant difference between the double mutant and the wild type in embryo sac area measurement ($p = 0.064$). To confirm these data, individual seed weights and abortion frequencies were measured (Table 6.2). The results showed that the final seed weight of the double mutant plant was not significantly different to the wild type (Col-0), and that there was no seed abortion ($p = 0.295$).

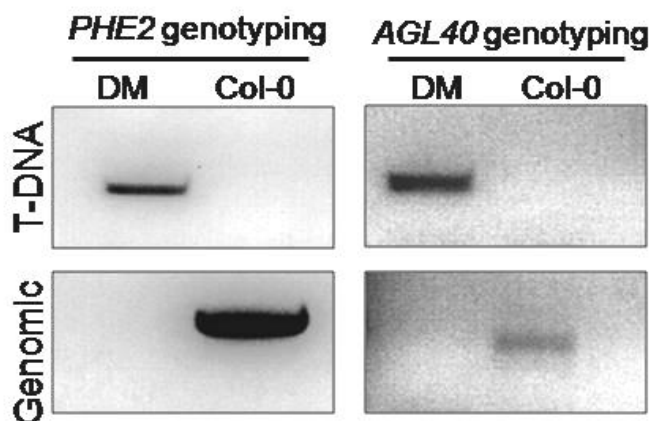


Figure 6.13 Genotyping of *phe2::agl40* homozygous double mutant F2 plants. The homozygous plant only showed a band in T-DNA PCR; the wild type (Col-0) only showed a band in genomic PCR. DM (double mutant); PCRs were performed on the same DNA samples.

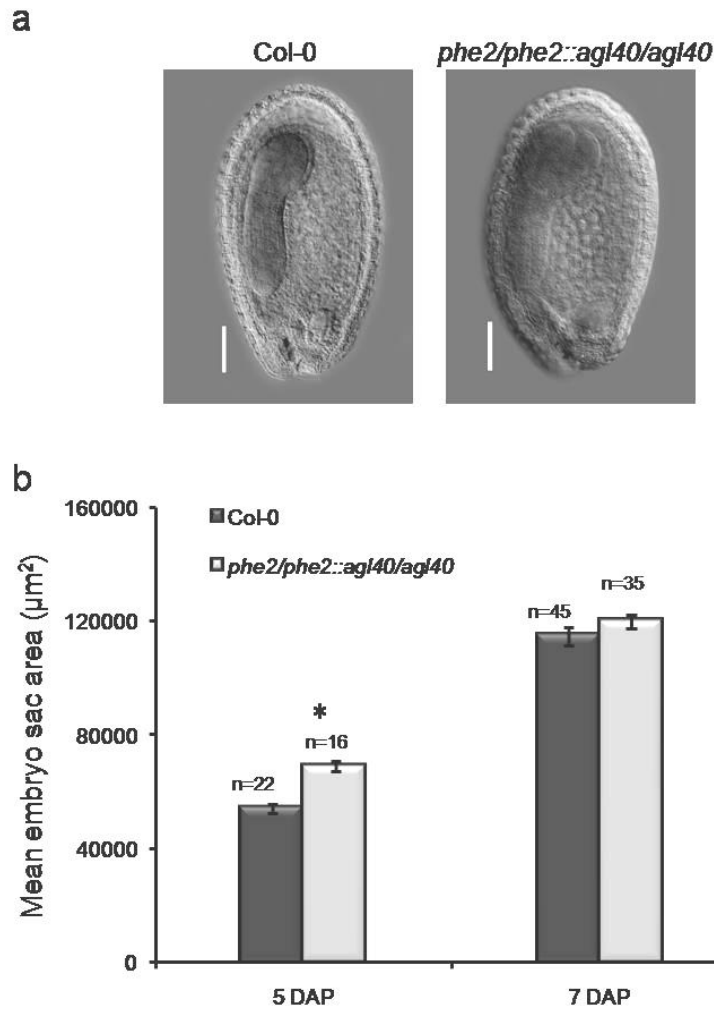


Figure 6.14 Seed phenotype and embryo sac area of wild type and *phe2/phe2::agl40/agl40* plants. (a) Cleared, developing seeds of Col-0 and the double mutant at 7 DAP were imaged using DIC optics. Bar size = 100 µm. (b) The columns represent the mean areas of the wild type (Col-0) and the double homozygous mutant (*phe2::agl40*) at 5 and 7 DAP. For comparisons, the value which differed at the 0.05 significance level between the wild type and the double mutant is labelled with *. Error bar = S.E.M.; n = number of seeds examined.

Table 6.2 Seed weight and abortion frequency of *phe2/phe2::agl40/agl40* and wild type plants.

Plant name	Mean seed weight (µg) ± S.E.M ^a	Number of seed	Average seed number/pod	p-value	Abortion%
Col-0	18.1 ± 0.47	152	44		0.0
<i>phe2/phe2::agl40/agl40</i>	17.3 ± 0.48	310	35	0.295 ^b	0.0

^a S.E.M seed weight of n ≥ 6 siliques.

^b Student's *t*-test with Col-0 wild type, at the 0.05 significant difference level.

6.2.4.2 Seed size analyses of *phe1/phe1::agl40/agl40* plants

The *phe1/phe1* single mutant affects endosperm proliferation, resulting in smaller seeds than the wild type (Chapter 4). On the other hand, *agl40/agl40* had no effect on seed size as a single mutant. In order to test the impact on seed development of the double mutant combining *phe1/phe1* and *agl40/agl40*, a cross was made between *phe1/phe1* and *agl40/agl40* plants to create F1 hybrid plants. Following this, F2 plants were genotyped to isolate the double homozygous *phe1/phe1::agl40/agl40* (Figure 6.15). To provide suitable control material, a cross was made between Col-0 and *Ler* wild types because the mutants were carried in different ecotypes: *phe1/phe1* from the *Ler* background and *agl40/agl40* from the Col-0 background. Seed size analysis was carried out on double homozygous plants by means of embryo sac area measurement and seed weight. The developing seeds of *phe1/phe1::agl40/agl40* were cleared and the embryo sac area sizes were measured at 5 and 7 DAP (Figure 6.16). The embryo sac area size of the *phe1/phe1::agl40/agl40* double mutant was significantly smaller than the wild type at 5 DAP ($p = 0.007$), whereas at 7 DAP no significant difference was observed between the area size of the double mutant and its wild type ($p=0.934$). In addition, seed weights were taken for both the wild type and the *phe1/phe1::agl40/agl40* double mutant seeds to test the effect of the double mutants on final seed size. Table 6.3 shows that the double homozygous mutant produced seeds 7% lighter than wild type seeds ($p = 0.017$).

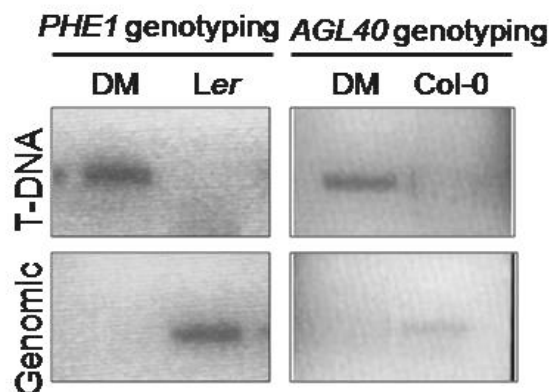


Figure 6.15 Genotyping of *phe1::agl40* homozygous double mutant F2 plants. The homozygous plant only showed a band in T-DNA PCR; the wild type (Col-0 or *Ler*) only showed a band in genomic PCR. DM (double mutant); PCRs were performed on the same DNA samples.

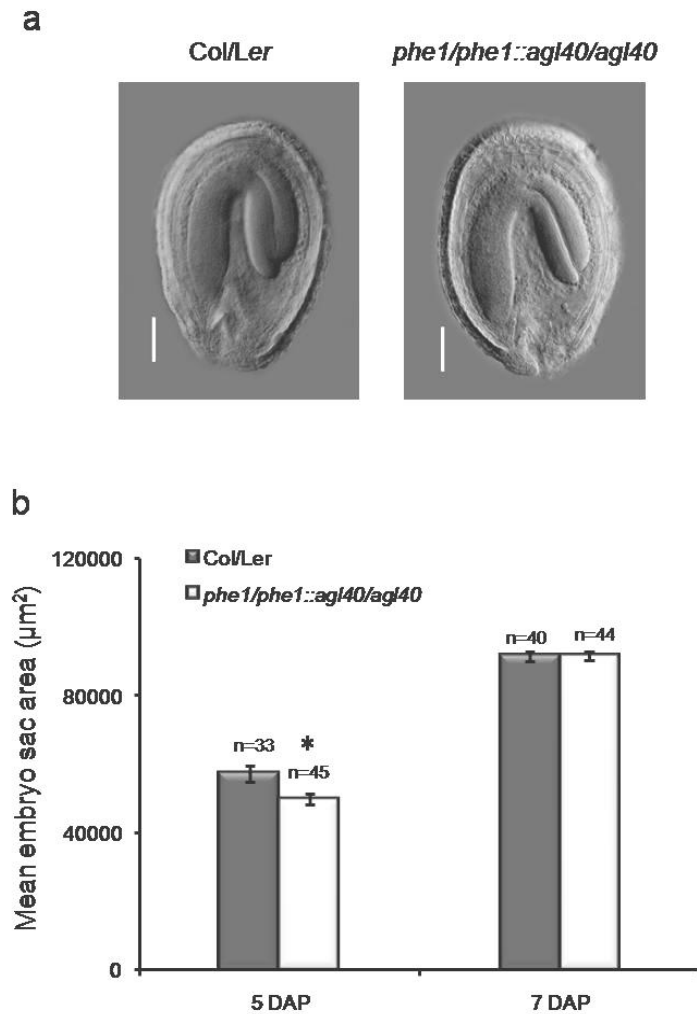


Figure 6.16 Seed phenotype and embryo sac area of wild type and *phe1/phe1::agl40/agl40* plants. (a) Cleared, developing seeds of the wild type (Col/Ler) and the double mutant at 7 DAP were imaged using DIC optics. Bar size = 100 μm. (b) The columns represent the mean area of the wild type (Col/Ler) and the double homozygous mutant (*phe1::agl40*) at 5 and 7 DAP. For comparison, the value which differed at the 0.05 significance level between the wild type and the double homozygous mutant is labelled with *. Error bar = S.E.M.; n = number of seeds examined.

Table 6.3 Seed weight and abortion frequency of *phe1/phe1::agl40/agl40* and wild type plants.

Genotype	Mean seed weight (μg) ± S.E.M ^a	Number of seed	Average seed number/pod	p-value	Abortion%
Col-0/Ler	18.5 ± 0.33	412	39		0.0
<i>phe1/phe1::agl40/agl40</i>	17.1 ± 0.40	720	43	0.017 ^{b*}	0.0

^a S.E.M seed weight of n ≥ 10 siliques

^b Student's *t*-test with Col-0/Ler wild type, at the 0.05 significant difference level. *Significantly different compared to wild type.

6.2.5 Analysis of triple mutants

There were no detectable differences in seed phenotype between the double homozygous mutant seeds and the wild type seeds. Since there are documented examples of double mutant combinations failing to produce mutant phenotypes, whilst triple mutants have strong impacts on development (e.g. *sep* mutants: Pelaz *et al.*, 2000), the next logical step in this work was to generate triple mutant combinations. Two parallel triple homozygous mutant combinations were created, one centred on the *phe2.1* double homozygous complex, and the other on the *phe1.1* double homozygous complex (Figure 6.17), in order to test their effects on endosperm development and final seed size.

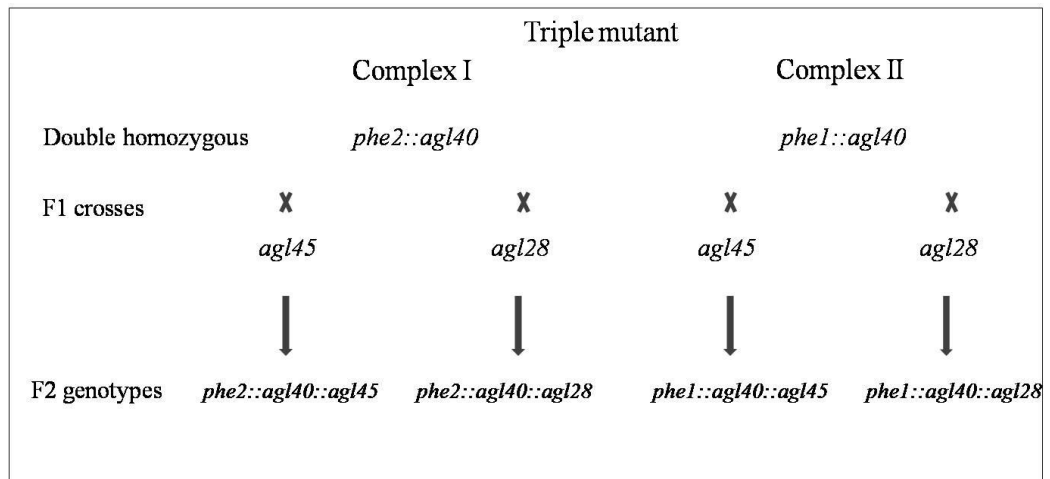


Figure 6.17 Diagram of the hypothetical triple mutant combinations centred around *phe2* and *phe1* homozygous mutants.

6.2.5.1 *phe2/phe2* triple homozygous complexes

There were four candidate genes (*AGL28*, *40*, *45* and *62*) that could be combined into triple mutant combinations with *phe2*. To minimise time and effort the strategy adopted was to add one gene at a time to the existing double mutant combination. In the *phe2* triple complexes, the first complex was produced by crossing *phe2/phe2::agl40/agl40* double mutant flowers with the *agl28/agl28* single mutant pollen to create an F1 heterozygote with a genotype of *PHE2/phe2::AGL40/agl40::AGL28/agl28*. Likewise, plants with the genotype *PHE2/phe2::AGL40/agl40::AGL45/agl45* were made by crossing *phe2/phe2::agl40/agl40* flowers with a *agl45/agl45* pollen parent. Both F1 plants of the triple complexes were segregated into F2s and genotyped to identify the triple homozygous

mutant plants (Figure 6.18). Approximately 60 F2 plants were analysed and two triple homozygous mutant plants were identified and kept for further analysis.

While screening and genotyping the F2 plants to find the triple homozygous mutants, we came across other double homozygous combinations that did not have *phe2.1* or *phe1.1*, such as *agl40::agl45* and *agl40::agl28*. In addition, these two double homozygous complexes had an overall wild-type phenotype.

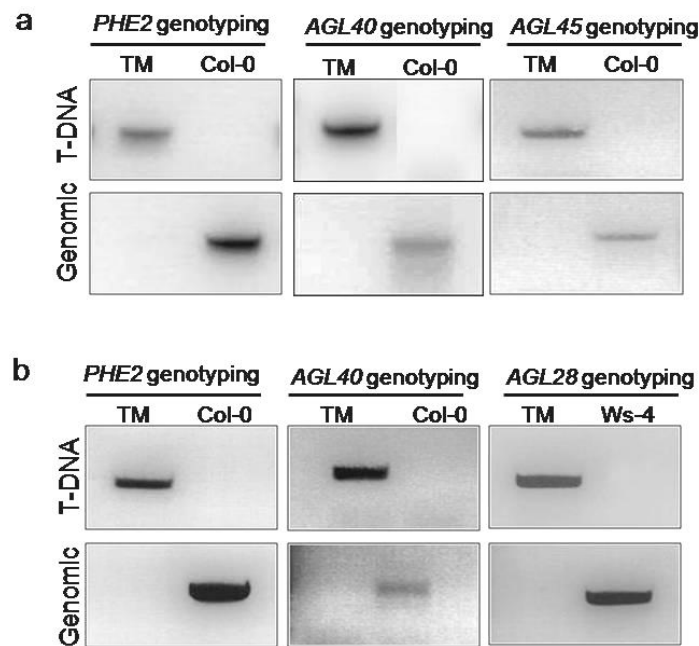


Figure 6.18 Genotyping of *phe2* triple homozygous F2 plants. The homozygous plants only showed bands in T-DNA PCR; the wild type plants only showed bands in genomic PCR. (a) A *phe2::agl40::agl45* homozygous plant. (b) A *phe2::agl40::agl28* homozygous plant. TM (triple mutant); Col-0 and Ws-4 were wild types; PCRs were performed on the same DNA samples for each plant.

6.2.5.1.1 Seed size analyses of *phe2/phe2::agl40/agl40::agl28/agl28* plants

The previous analysis of the *phe2* double homozygous complex (*phe2/phe2::agl40/agl40*) had revealed no significant differences in embryo sac area size and final seed weight when compared to the wild type. In order to test the effect of the *phe2* triple complex (*phe2/phe2::agl40/agl40::agl28/agl28*) on endosperm proliferation and final seed size, embryo sac area and seed weight measurements were carried out on the triple mutant and the corresponding wild types. The developing, cleared seeds of the triple mutant and wild

types at 3 and 7 DAP showed no obvious changes in seed phenotype (Figure 6.19a). The embryo sac area size of the triple homozygous mutant (*phe2/phe2::agl40/agl40::agl28/agl28*) was not significantly different compared to the Col-0 wild type at 3 DAP ($p = 0.148$), but it was 26% smaller compared to the Ws-4 wild type ($p < 0.001$). Conversely, at 7 DAP, the triple mutant embryo sac area was not significantly different compared to the Ws-4 wild type ($p = 0.378$), but it was significantly smaller (9.4% decrease, $p = 0.002$) than the Col-0 wild type (Figure 6.19b). The final seed weight of the *phe2/phe2::agl40/agl40::agl28/agl28* plants was not significantly different compared to the wild type plants (Col-0, $p = 0.177$; Ws-4, $p = 0.913$) (Table 6.4).

6.2.5.1.2 Seed size analyses of *phe2/phe2::agl40/agl40::agl45/agl45* plants

The second *phe2* triple homozygous complex tested was *phe2/phe2::agl40/agl40::agl45/agl45*. Figure 6.20a shows that the developing seed of *phe2/phe2::agl40/agl40::agl45/agl45* had a similar seed morphology and size to the wild type Col-0 at 3 DAP. However, at 7 DAP embryo development in the mutant was further advanced compared to the wild type. The embryo sac area size of the *phe2/phe2::agl40/agl40::agl45/agl45* seeds was not significantly different compared to the wild type (Col-0) at 3 DAP ($p=0.194$). At 7 DAP the triple mutant seeds were slightly larger than Col-0 seeds (10.7%, $p=0.001$); this correlated with the relatively advanced embryo development of the mutant seed compared to the wild type at 7 DAP (Figure 6.20b). These results confirmed that there was no significant difference in seed weight between *phe2/phe2::agl40/agl40::agl45/agl45* and the Col-0 wild type ($p = 0.672$; Table 6.4).

In summary, the seed weight data showed that the two *phe2* triple mutant complexes *phe2/phe2::agl40/agl40::agl28/agl28* and *phe2/phe2::agl40/agl40::agl45/agl45* had no effect on seed development as measured by embryo sac area and final seed weight.

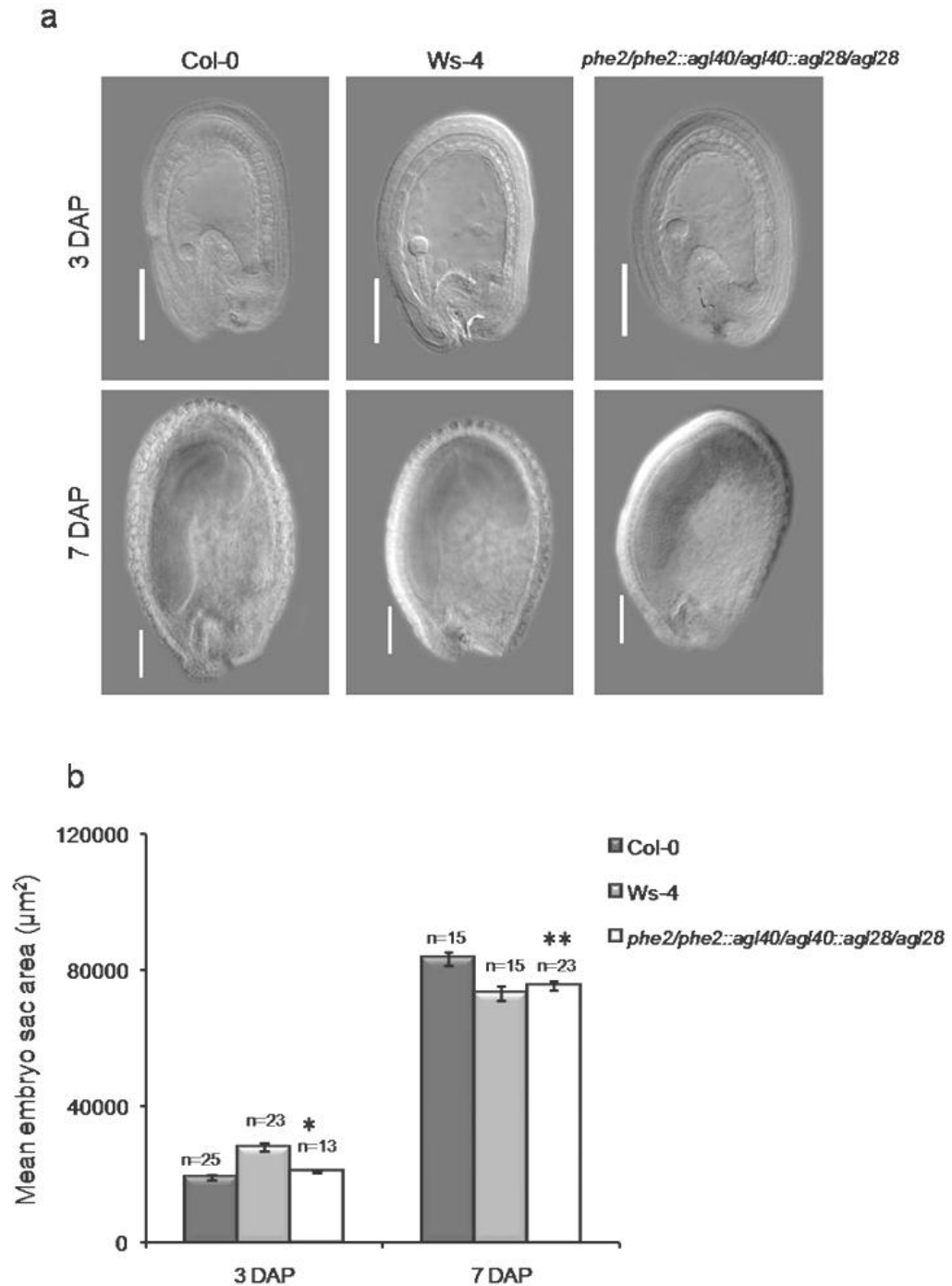


Figure 6.19 Seed phenotype and embryo sac area of wild type and *phe2/phe2::agl40/agl40::agl28/agl28* plants. (a) Cleared, developing seeds of the wild types (Col-0 and Ws-4) and the triple mutant at 3 and 7 DAP were imaged using DIC optics. Bar size = 100 μm . (b) The columns represent the mean areas of the wild types Col-0 and Ws-4 and the *phe2/phe2::agl40/agl40::agl28/agl28* at 3 and 7 DAP. For comparison, the values which differed at the 0.05 significance level between the wild type Ws-4 and the triple mutant are labelled with * and those between the wild type Col-0 and the triple mutant are labelled with **. Error bar = S.E.M.; n = number of seeds examined.

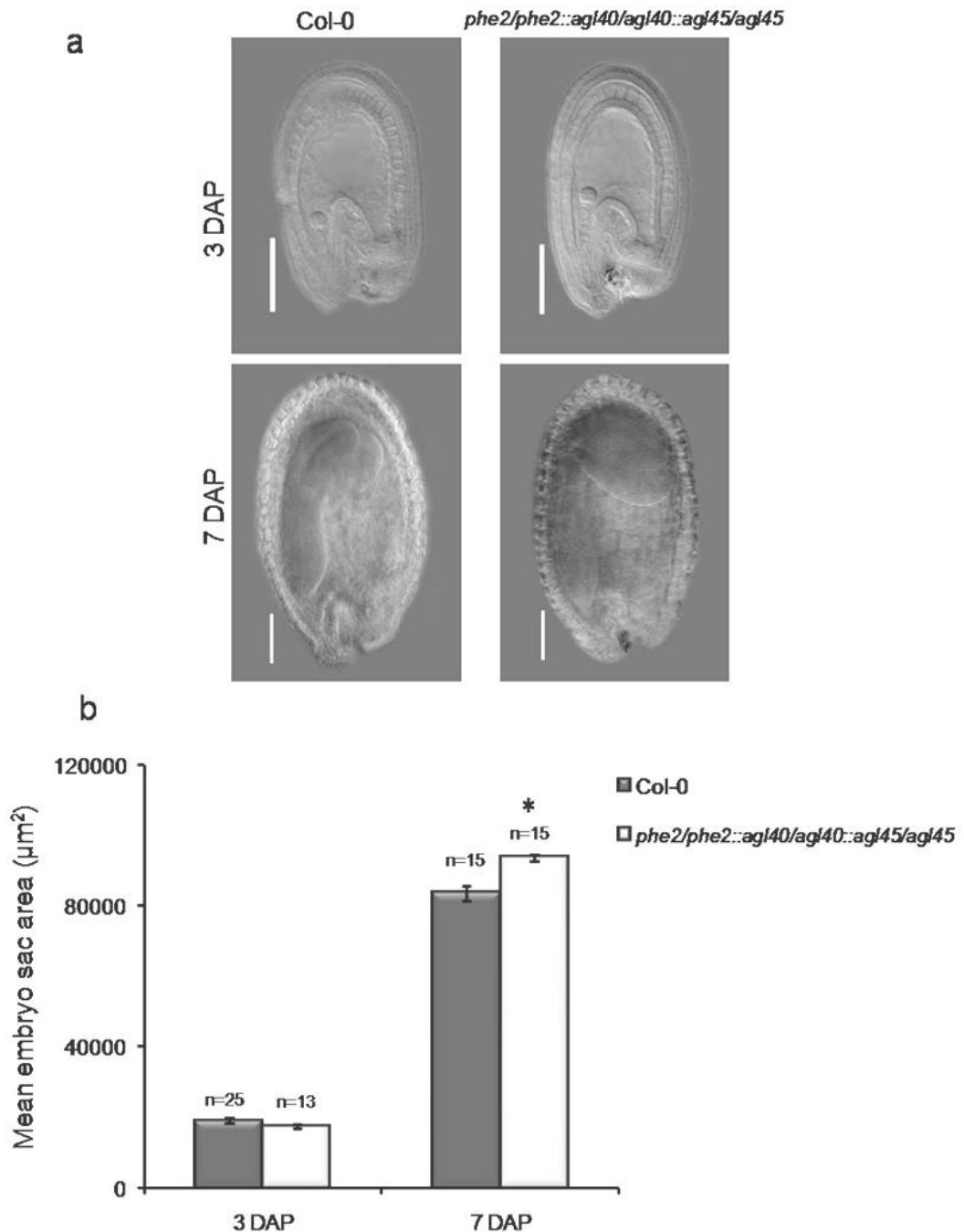


Figure 6.20 Seed phenotype and embryo sac area of wild-type and *phe2/phe2::agl40/agl40::agl45/agl45* plants. (a) Cleared, developing seeds of the wild type Col-0 and the triple mutant at 3 and 7 DAP were imaged using DIC optics. Bar size = 100 μm . (b) The columns represent the mean areas of the wild type Col-0 and the *phe2/phe2::agl40/agl40::agl45/agl45* at 3 and 7 DAP. For comparison, the value which differed at the 0.05 significance level between the wild type and the triple mutant is labelled with *. Error bar = S.E.M.; n = number of seeds examined.

Table 6.4 Seed weight and abortion frequency of the *phe2* triple homozygous complexes and the wild types.

Genotype	Seed weight (μg) \pm S.E.M ^a	Number of seed	Average number of seed/pod	p-value	Abortion%
Col-0	18.8 \pm 0.33	554	52		0.0
Ws-4	17.9 \pm 0.60	497	40		0.0
<i>phe2/phe2::agl40/agl40</i> <i>::agl28/agl28</i>	17.9 \pm 0.45	869	55	0.177 ^b 0.913 ^c	0.0
<i>phe2/phe2::agl40/agl40</i> <i>::agl45/agl45</i>	19.0 \pm 0.59	621	48	0.672 ^b	0.0

phe2 and *agl40* were from Col-0 genetic background, *agl28* from Ws-4 genetic background

^a S.E.M seed weight of $n \geq 10$ siliques.

^b Student's *t*-test with Col-0 wild type, at the 0.05 significant difference level.

^c Student's *t*-test with Ws-4 wild type, at the 0.05 significant difference level.

6.2.5.2 *phe1/phe1* triple homozygous complexes

Triple mutants centred on *phe1* were generated using the same approach as described above for *phe2*. The first complex was produced by crossing *phe1/phe1::agl40/agl40* double mutant flowers with the *agl28/agl28* single mutant pollen to create F1 hybrid heterozygote plants with a genotype of *PHE1/phe1::AGL40/agl40::AGL28/agl28*. The second *phe1* complex was made by crossing *phe1/phe1::agl40/agl40* flowers with a single homozygous *agl45/agl45* pollen to produce *PHE1/phe1::AGL40/agl40::AGL45/agl45* F1 heterozygous seeds. Both F1 plants selfed to produce segregating F2s, from which the triple mutants were isolated. Genotyping ~58 *PHE1/phe1::AGL40/agl40::AGL28/agl28* F2 plants (Figure 6.21a) identified two triple mutant plants that were kept for further analyses. In contrast, genotyping more than sixty F2 plants originating from the F1 plant with a genotype of *PHE1/phe1::AGL40/agl40::AGL45/agl45* failed to identify any triple homozygous mutants. Therefore, a plant with a genotype of *phe1/phe1::agl40/agl40::AGL45/agl45* was kept for further analyses (Figure 6.21b).

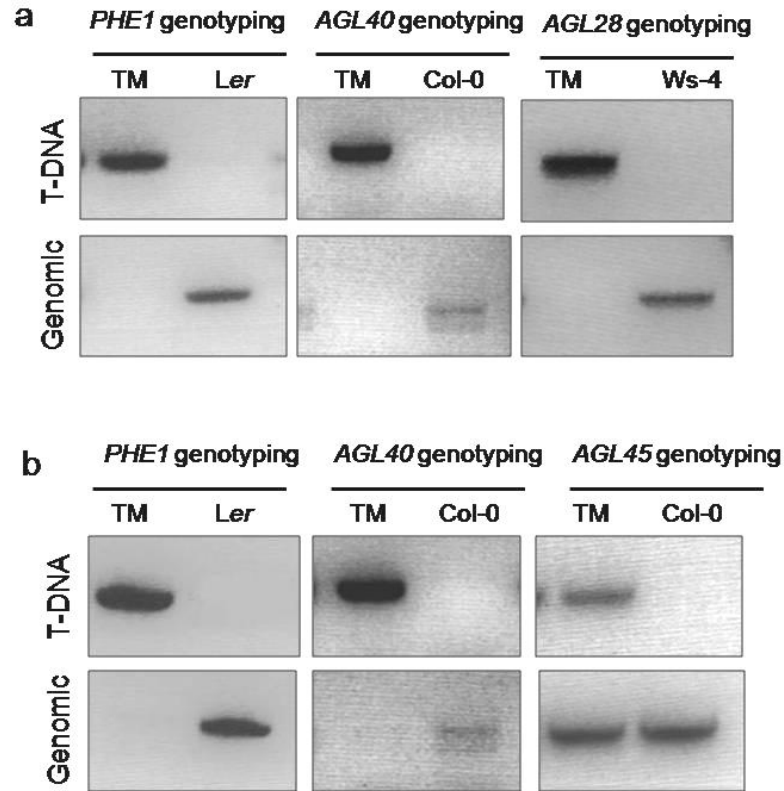


Figure 6.21 Genotyping of *phe1* triple homozygous F2 plants. The homozygous plants only showed bands in T-DNA PCR; the wild type only showed bands in genomic PCR. (a) A *phe1::agl40::agl28* homozygous plant. (b) A *phe1/phe1::agl40/agl40::AGL45/agl45* plant. TM (triple mutant); Col-0, *Ler* and *Ws-4* were wild types; PCRs were performed on the same DNA samples of each plant.

6.2.5.2.1 Seed size analyses of *phe1/phe1::agl40/agl40::agl28/agl28* plants

The *phe1/phe1::agl40/agl40* double homozygous complex was not significantly different compared to the wild type in embryo sac area size and only slightly differed in final seed weight. The triple mutant *phe1/phe1::agl40/agl40::agl28/agl28* was therefore made using the approach outlined in the previous section. Endosperm proliferation, final seed size, embryo sac area and seed weight measurements were carried out on the triple mutant and wild types. The developing, cleared seeds of the triple mutant were similar in seed phenotype compared to the wild types (*Col-0* and *Ler*) at 3 DAP, but they were smaller than the *Ws-4* wild type. However, at 7 DAP, the triple mutant was similar in seed phenotypes to the *Col-0* and *Ws-4* wild types but larger than the *Ler* wild type (Figure 6.22a). Although the embryo sac area size of the triple homozygous mutant was not significantly different compared to *Col-0* ($p = 0.919$) and *Ler* ($p = 0.066$) wild types at 3 DAP, there was a significantly smaller embryo sac area size (32.5% decrease; $p < 0.000$) when compared to the *Ws-4* wild type. In contrast, at 7 DAP, the triple mutant embryo sac

area size was not significantly different compared to the Col-0 ($p = 0.052$) and Ws-4 ($p = 0.138$) wild types, but it was 23% ($p < 0.000$) larger than the *Ler* wild type (Figure 6.22b). However, the mean weight of the seeds harvested from triple mutant plants was not significantly different compared to the wild types: Col-0 ($p = 0.967$), *Ler* ($p = 0.360$) and Ws-4 ($p = 0.434$) (Table 6.5).

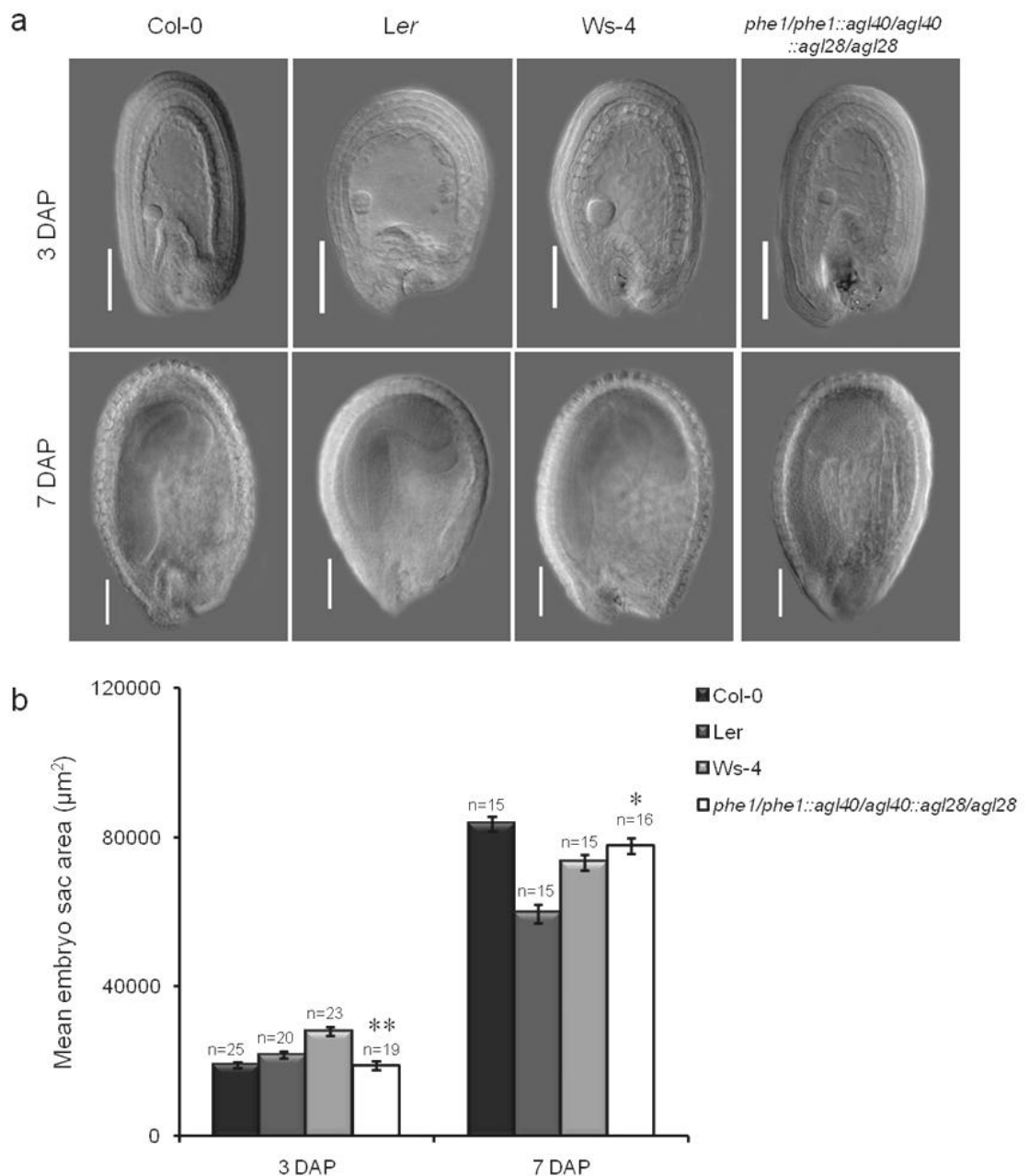


Figure 6.22 Phenotype and embryo sac area of wild type and *phe1/phe1::agl40/agl40::agl28/agl28* seeds. (a) Cleared, developing seeds of the wild types (Col-0, *Ler* and Ws-4) and the triple mutant at 3 and 7 DAP were imaged using differential interference contrast microscopy. Bar size = 100 µm. (b) Embryo sac areas at 3 and 7 DAP. For comparison, values which differed at the 0.05 significance level between the wild type *Ler* and the triple mutant are labelled with * and those between the wild type (Ws-4) and the triple mutant are labelled with **. Error bar = S.E.M.; n = number of seeds examined.

Table 6.5 Seed weight and abortion frequency of the *phe1* triple homozygous complex and the wild types.

Genotype	Mean seed weight (μg) \pm S.E.M ^a	Number of seed	Average seed number/pod	p-value	Abortion%
Col-0	18.8 \pm 0.33	554	52		0.0
Ler	18.2 \pm 0.40	760	41		0.0
Ws-4	17.9 \pm 0.60	497	40		0.0
<i>phe1/phe1::agl40/agl40::agl28/agl28</i>	18.7 \pm 0.56	354	45	0.967 ^a 0.360 ^b 0.434 ^c	0.0

phe1, *agl40* and *agl28* were from the Ler, Col-0 and Ws-4 genetic backgrounds, respectively.

^a S.E.M seed weight of $n \geq 8$ siliques.

^b Student's *t*-test with Col-0 wild type, at the 0.05 significant difference level.

^c Student's *t*-test with Ler wild type, at the 0.05 significant difference level.

^d Student's *t*-test with Ws-4 wild type, at the 0.05 significant difference level.

6.2.5.2.2 Seed phenotype analyses of *phe1/phe1::agl40/agl40::AGL45/agl45* plants

As described in Section 6.2.5.2, the genotyping of approximately sixty F2 plants originating from a *PHE1/phe1::AGL40/agl40::AGL45/agl45* F1 plants failed to identify any *phe1/phe1::agl40/agl40::agl45/agl45* triple mutants. Therefore, a closely related genotype, *phe1/phe1::agl40/agl40::AGL45/agl45*, was chosen for further analysis with the objective of understanding the failure to recover triple mutant seeds.

The selected mutant plant was vegetatively similar to the wild type. However, young siliques (6-14 DAP) of selfed *phe1/phe1::agl40/agl40::AGL45/agl45* plants contained a mixture of highly abnormal seeds and seeds indistinguishable from the Col-0 wild type. At 6-8 DAP the silique contained normal green seeds and large white seeds; however, at 9-14 DAP the silique contained brown collapsed seeds as well as normal green seeds (Figure 6.23a and b). To investigate this phenotype further, a mean seed set percentage was measured for both the normal green seeds and the abnormal brown seeds that presented in the same silique (30 siliques at 12-14 DAP) and it was observed that there were $75\% \pm 2.26$ normal green seeds and $25\% \pm 2.26$ abnormal brown collapsed seeds ($p < 0.001$) (Figure 6.23m). The abortion frequency of about 25% is consistent with the proposition that triple *phe1::agl40::agl45* mutants cause seed lethality. Seeds cleared at the late-embryogenesis stage (8-9 DAP) contained embryos of both a wild-type appearance (Figure 6.23c) and abnormally small, under-developed and/or misshapen embryos (Figure 6.23d-k). In addition, the embryo sac area of the abnormal and collapsed seed was significantly

smaller (58%; $p < 0.001$) than the normal plump seed from the same silique (Figure 6.23I). This data suggests that seed abortion of a proportion of F2 seeds is associated with the cessation of embryo development. The abnormal seed subjected to a germination test failed to germinate, suggesting they were unviable. In summary, although the embryos in the abnormal seeds found in selfed pods of *phe1/phe1::agl40/agl40::AGL45/agl45* plants could not be genotyped, a reasonable assumption was that these seeds represented the missing *phe1/phe1::agl40/agl40::agl45/agl45* genotype. Unfortunately, the work reported in this section had to stop here due to lack of time.

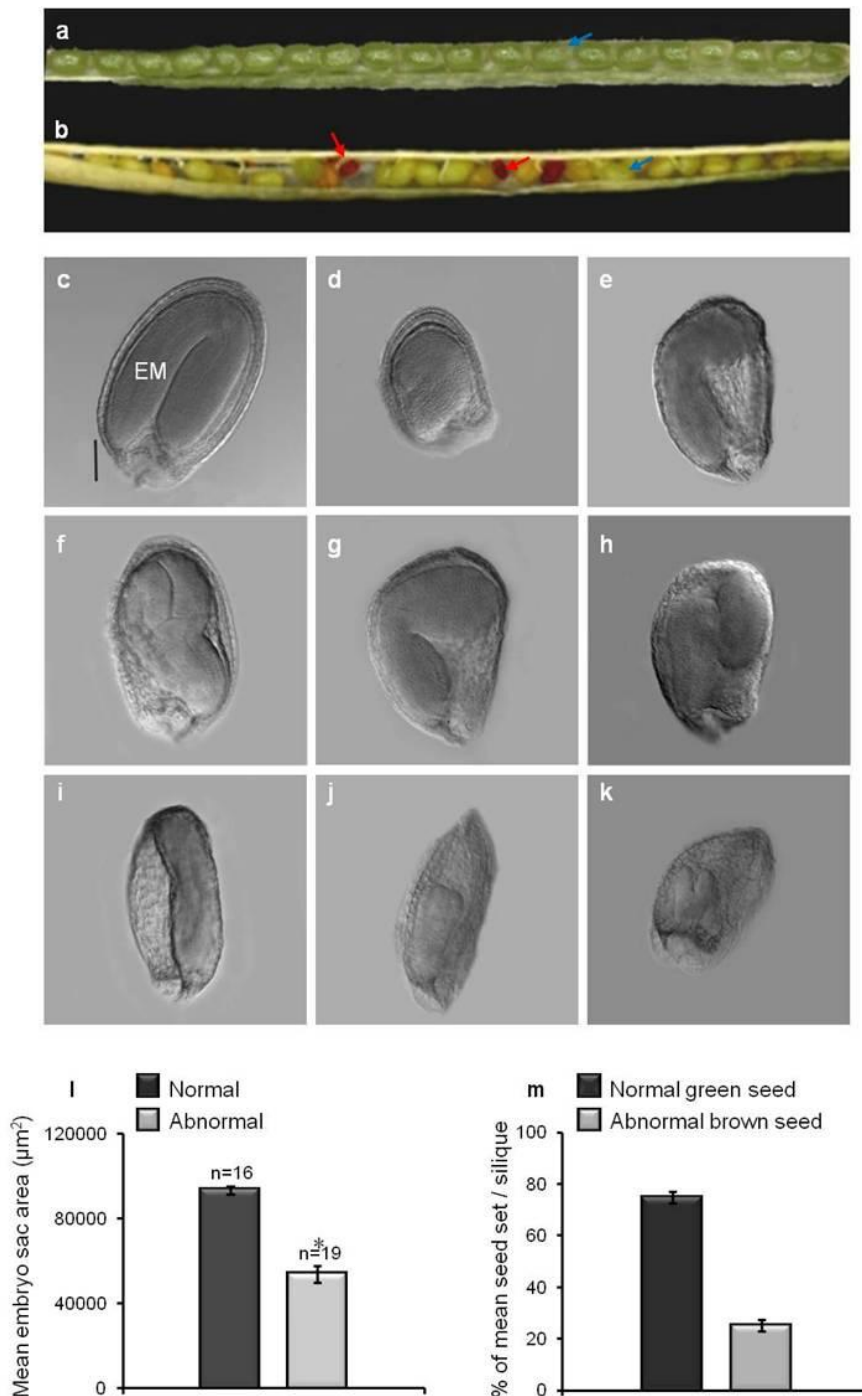


Figure 6.23 Seed phenotype of the *phe1/phe1::agl40/agl40::AGL45/agl45* mutant and the wild type. a-b, Seeds present in a young silique (12 DAP) of (a) Col-0, and (b) a *phe1/phe1::agl40/agl40::AGL45/agl45* mutant showing a normal green seed (blue arrow) and a brown collapsed seed (red arrow). c-k, Cleared seeds at the mature embryo developmental stage of the *phe1/phe1::agl40/agl40::AGL45/agl45* mutant plant (8-9 DAP), showing normal seeds with a mature cotyledon embryo (c) and collapsed seeds with a ceased embryo, d-k; bar size = 100 μm. (l) Embryo sac area measurement of normal seeds and abnormal collapsed seeds. The columns represent the mean embryo sac area sizes; n = number of seeds examined. (m) The percentages of normal and abnormal seed sets per young silique (12-14 DAP) of the *phe1/phe1::agl40/agl40::AGL45/agl45* mutant plant. The columns represent the mean seed set percentages of (30 siliques). For comparison, the value which differed at the 0.05 significance level between the normal seeds and the brown collapsed seeds was labelled with *. Error bar = S.E.M.

6.2.6 *phe2*-centered quadruple mutant

Seed analyses of the two *phe2* triple mutant complexes discussed previously (Section 6.2.5.1) indicated that there was no effect of the triple mutants (*phe2/phe2::agl40/agl40::agl28/agl28* and *phe2/phe2::agl40/agl40::agl45/agl45*) on seed phenotype, especially on final seed weight. Therefore, the *phe2* quadruple mutant complex was created to test the effect on seed development of combining all four mutants *phe2*, *agl40*, *agl45* and *agl28*. The quadruple mutant was made by crossing a triple homozygous flower (*phe2/phe2::agl40/agl40::agl28/agl28*) with a single homozygous pollen (*agl45/agl45*) to produce F1 seeds and plants with a genotype of (*PHE2/phe2::AGL40/agl40::AGL28/agl28::AGL45/agl45*). The F1 plants were selfed to give an F2 generation and a subsequent F3 generation. However, extensive genotyping of both F2 (~100 plants) and F3 (~80 plants) plants failed to produce a single quadruple homozygous plant. A closely related genotype (*phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28*) (Figure 6.24) was therefore selected for further seed phenotype analyses. In addition, while genotyping the *phe2* quadruple complex we came across a genotype (*agl40/agl40::agl45/agl45::agl28/agl28*) of a triple homozygous plant which showed an overall phenotype similar to the wild type and no seed abortion was detected.

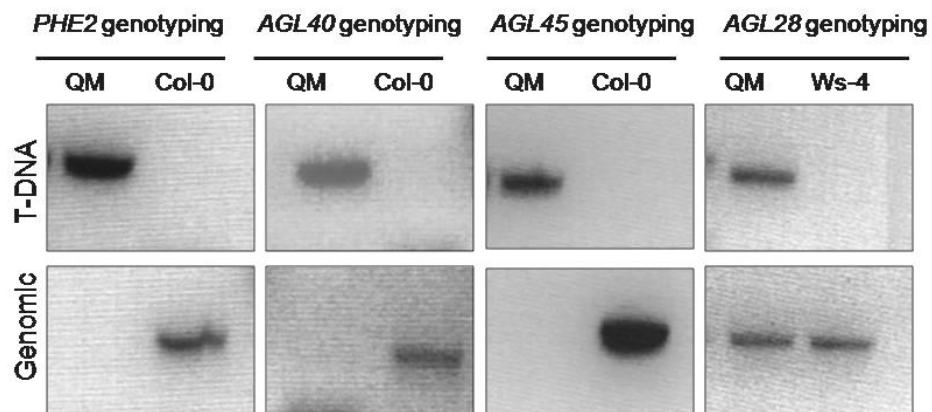


Figure 6.24 Genotypes of the *phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28* F3 plant. Plants homozygous for a particular insertion only produced a product in T-DNA PCR; plants with a wild type allele only produced a PCR product in genomic PCR. QM (potential quadruple mutant); Col-0, Ler and Ws-4 were wild types; PCRs were performed on the same DNA sample of each plant.

6.2.6.1 Seed phenotype analyses of *phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28* plants

The quadruple mutant plant was similar in vegetative phenotype to the *phe2* triple homozygous complexes and the wild type; however, the developing seeds in the young silique were different. The *phe2* quadruple mutant (*phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28*) plant showed an abnormal seed phenotype, with pods containing a mixture of seeds with a normal appearance and white or small shrivelled seeds which were absent from the *phe2* triple homozygous plants with genotypes of (*phe2/phe2::agl40/agl40::agl28/agl28* and *phe2/phe2::agl40/agl40::agl45/agl45*) (Figure 6.25a). A developing white seed and a normal seed were collected and cleared at 3, 5, 7 and 9 DAP in order to visualise endosperm and embryo development. Figure 6.26 shows a comparison between the normal and white seeds from the *phe2* quadruple mutant plant. The white seed was smaller in size and contained an unstructured endosperm with no material or cells and an abnormal irregular shaped embryo at all of the tested developmental stages. In order to confirm the result of seed clearing, the abortion percentage was measured for the *phe2* quadruple mutant plant and compared with the two *phe2* triple homozygous plants. The abortion percentage of the *phe2* quadruple mutant plant was $25.3\% \pm 0.59$ ($p < 0.001$) (Figure 25b-c), whereas in both *phe2* triple complexes there was no abortion (see also Section 6.2.5.1).

Although direct genotyping of the abnormal white and shrivelled seeds was not possible, the 75%:25% ratio of viable:aborted seeds suggests these likely represent the quadruple mutant (*phe2/phe2::agl40/agl40::agl45/agl45::agl28/agl28*) progeny that could not be recovered from selfed *phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28* plants (Section 6.2.6). This indicates that the quadruple mutant has a lethal effect on seed development and therefore that this complex plays a crucial role.

In order to investigate how the *phe2/agl40/agl45/agl28* homozygous mutant combination causes seed abortion, a confocal microscopy investigation was conducted on seeds from *phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28* selfed siliques across a range of developmental stages: 3, 5, 7 and 9 DAP (Figure 6.27). The abnormal seeds (Figure 6.27m-p) were smaller than the normal seed (Figure 6.27a-d) at all stages and exhibited conspicuous abnormalities in both embryo and endosperm development from the earliest time point tested (3 DAP). The embryo consisted of an irregularly shaped cluster of cells of a similar size and morphology in seeds at all stages, indicating that embryo development

was halted at an early stage. The endosperm of the normal seed showed an entirely wild-type pattern of development (Figure 6.27i-l), whereas in the abnormal seed the embryo sac contained an amorphous material and no normal endosperm nuclei or cells (Figure 6.27u-x). The relative severity of the endosperm phenotype compared to that of the embryo suggests that the primary reason for seed abortion was a failure in endosperm development. In this case, the abnormal appearance of the embryo and its failure to progress would be secondary consequences of endosperm failure.

In summary, the data presented above strongly suggest that the quadruple mutant *phe2/phe2::agl40/agl40::agl45/agl45::agl28/agl28* has a lethal effect on seed development, with the primary effect potentially within the endosperm.

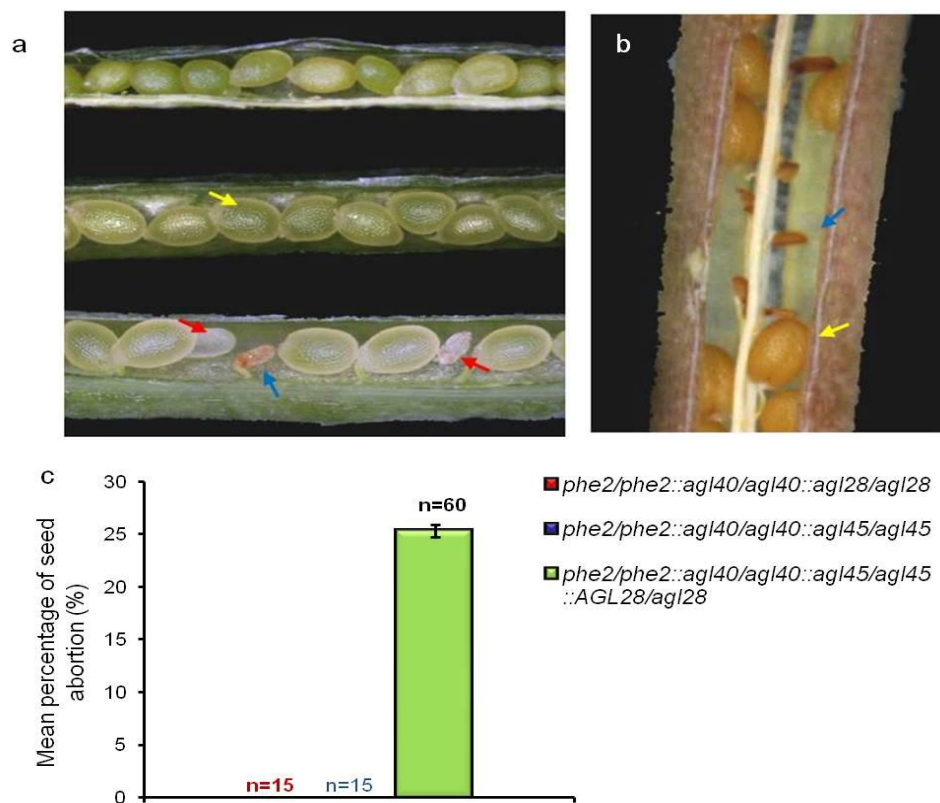


Figure 6.25 Seed phenotype and abortion percentage of *phe2* quadruple complex compared to the *phe2* triple complex. (a) Seeds in the silique of the *phe2/phe2::agl40/agl40::agl28/agl28* mutant (top, ×4), the *phe2/phe2::agl40/agl40::agl45/agl45* mutant (middle, ×4) and the *phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28* mutant (bottom, ×5); yellow arrow = normal viable seed, red arrow = abnormal white seed and blue arrow = shrivelled seed. (b) The *phe2* quadruple complex mature seeds in a dry silique showed shrivelled (blue arrow) and viable (yellow arrow) seeds. (c) Graph showing seed abortion percentage of the *phe2* quadruple mutant (25%) compared with both *phe2* triple homozygous complexes (0%), n = number of siliques.

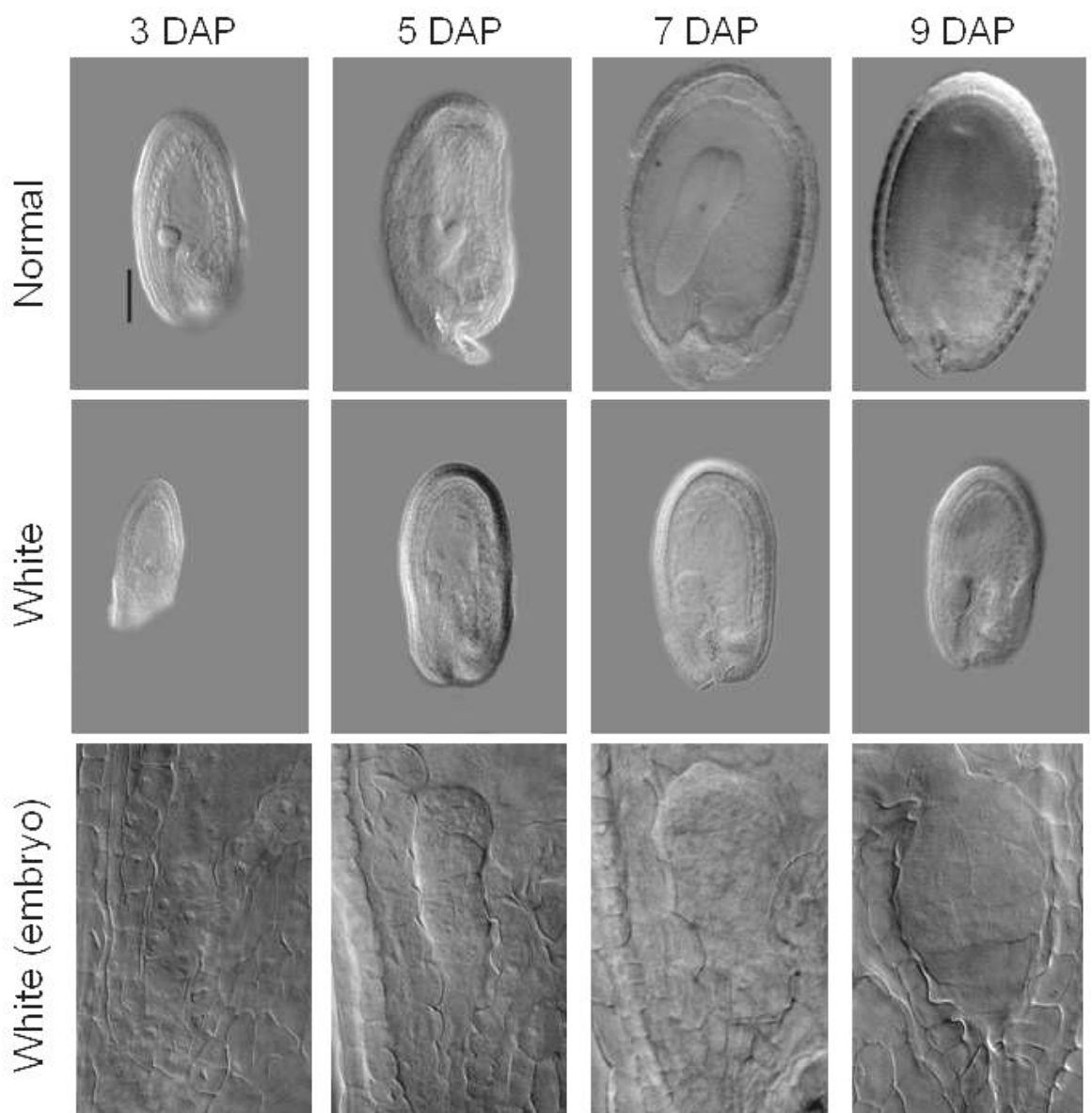


Figure 6.26 White seed phenotype of the *phe2* quadruple mutant compared to the normal seed at different DAP. Cleared, developing seeds at 3, 5, 7 and 9 DAP of the *phe1/phe1::agl40/agl40::agl45/agl45::AGL28/agl28* normal (top line); white seeds (middle line); embryo of the white seeds at a higher magnification from the same pictures in the middle line (bottom line). For the pictures in the top and middle lines: bar size = 100 μ m.

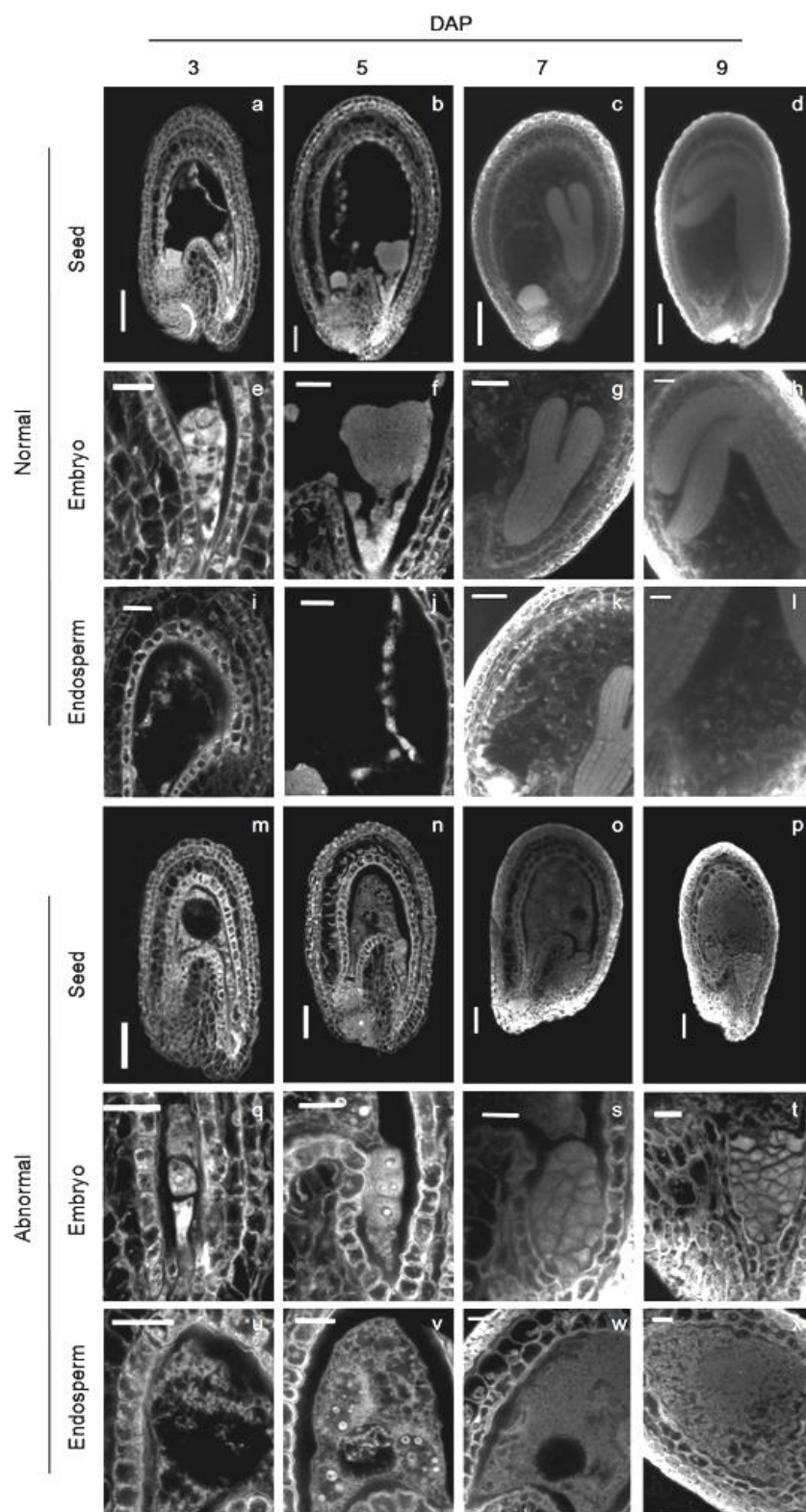


Figure 6.27 Confocal microscopy images of the *phe2* quadruple mutant seed at different days after pollination. a-d, developing normal seed at 3, 5, 7 and 9 DAP, respectively. e-h, developing normal embryo at 3, 5, 7 and 9 DAP, respectively. i-l, developing normal endosperm at 3, 5, 7 and 9 DAP, respectively. m-p, developing abnormal seed at 3, 5, 7 and 9 DAP, respectively. q-t, developing abnormal embryo at 3, 5, 7 and 9 DAP, respectively. u-x developing abnormal endosperm at 3, 5, 7 and 9 DAP, respectively. DAP = days after pollination. a, b, g, k, m-p, scale bar = 50 μ m; c and d, scale bar = 100 μ m; e, f, h, i, j, l, q-x, scale bar = 20 μ m.

6.3 Discussion

Our knock out strategy was designed to study the role of genes selected by transcriptional profiling as potentially important in endosperm development (Section 6.1.1). The candidate genes *PHE2*, *AGL40*, *AGL45* and *AGL28* were highly expressed in paternalized endosperm and belong to a MADS box family (Chapter 3; Tables 3.1, 3.2). The MADS box family of proteins are transcriptional regulators involved in numerous biological functions in plants (Alvarez-Buylla *et al.*, 2000a; Alvarez-Buylla *et al.*, 2000b; Parenicova *et al.*, 2003). The work reported in this chapter attempted to determine whether or not any single or combinations of loss-of-function T-DNA insertion alleles resulted in any detectable changes to endosperm development.

6.3.1 Homozygous T-DNA insertions did not affect the vegetative phase

Overall plant phenotypes such as stem length, leaf shape and size, number of branches, flower symmetry and flowering time of the *phe2.1*, *agl40.1*, *agl45.1* and *agl28.1* homozygous insertion lines were similar compared to the wild type, suggesting that expression of these genes is highly restricted or that they act redundantly. The transcriptional profiling revealed that the transcriptome of these genes was exclusively in the seed of *A. thaliana*, as reported by different databases (Chapter 3; Table 3.2), which indicated that the expression of these genes is restricted in the seeds. This data is therefore consistent with the absence of vegetative phenotypes being associated with the insertion alleles.

6.3.2 Endosperm size and individual seed weight

Experiments were carried out to measure the embryo sac area size of developing seeds and individual seed weight obtained from T-DNA insertion homozygous mutants compared to the wild type in order to characterize the effect of the candidate genes on endosperm development.

6.3.2.1 The *phe2.1*, *agl40.1*, *agl45.1* and *agl28.1* mutants have no discernable effect on seed development

The *agl40.1*, *agl45.1* and *agl28.1* homozygous insertion lines displayed no significant differences in embryo sac area size or individual seed weight compared to their wild types. Although the *phe2.1* homozygous mutant was not significantly different compared to its wild type (Col-0) in embryo sac area size, in contrast to the other mutants there was a small

(6%) but statistically significant increase in the weight of the mature seed. A possible explanation for this is that the embryo sac area size only measured the endosperm size but the final seed weight included contributions from the embryo, the endosperm and the seed coat (integument layer). Also, although there was a statistically significant difference in seed weight between the *phe2.1* mutant (20 µg) and the wild type Col-0 (19 µg), this variation cannot be considered as biologically significant.

One possible explanation for the near absence of seed phenotype associated with the single mutants is functional redundancy. This is frequently cited when the knock-out of a single gene gives no mutant phenotype, particularly when the gene belongs to a gene family (Bouche and Bouchez, 2001; Moore *et al.*, 2005). An example of functionally redundant genes which belong to the same MADS box family is *SEPALLATA1-3* (*SEP1-3*), which plays a role in the regulation of flower organ patterning and development (Pelaz *et al.*, 2000). Single *sep* mutants did not generate an observable phenotype, but in the triple mutant, *sep1/sep2/sep3*, all flower organs were transformed into sepals. Similarly, the MADS box genes *SHATTERPROOF1* and 2 (*SHP1-2*), which are involved in pod shattering, have no phenotypic effect singly, but prevent pod shatter when combined in the double mutant (*shp1/shp2*) (Liljegren *et al.*, 2000).

de Folter *et al.* (2005) carried out a study using a yeast two-hybrid system and provided a comprehensive protein-protein interactome map of the MADS box family in *A. thaliana*. The authors concluded that the control of biological processes depends on protein interactions, and completely redundant proteins are expected to have the same interaction patterns and be involved in the same developmental process. According to these interacting patterns, PHE1 interacted with AGL28, AGL40, and AGL62, and all showed coexpression with PHE1 in the embryo. Figure (6.1) illustrates the interaction pattern of all candidate genes in this present study: PHE2 interacts with AGL62, AGL40 and AGL28, and AGL28 interacts with PHE1 and PHE2. Moreover, AGL45 only interacts with AGL40, while AGL40 interacts with PHE1 and PHE2. Although little is known about the functions of *PHE2*, *AGL40*, *AGL45* and *AGL28* in seed development, *AGL62* and *PHE1* have a crucial role as growth promoters on seed development, as illustrated in previous chapters.

In summary, the proteins of our candidate genes in this study either interacted directly or indirectly in the yeast two-hybrid analysis, which may indicate that they have a similar function in plants, or that they combine to regulate a specific function or pathway in plants.

6.3.3 Double mutants do not impact on seed development

To begin to understand the relationship between the four candidate genes, experiments were carried out to create double homozygous mutants. The *PHE1* and *PHE2* gene loci are physically very close to each other on chromosome 1 (~12 kb apart); therefore, double mutants could not be made by a sexual route. Consequently, two ‘complexes’ were created, one with *phe1/phe1* and the other with the *phe2/phe2* homozygous mutant. The double homozygous mutants *phe2::agl40* and *phe1::agl40* were created and selected as double homozygous lines for further experiments. No phenotypic changes, such as stem length, leaf size, or flower size or symmetry were detected in double homozygous complex *phe2::agl40* and *phe1::agl40* compared to the wild type. This was also the case for the endosperm phenotype of the developing seeds and the final seed weight of *phe2/phe2::agl40/agl40* and *phe1/phe1::agl40/agl40*. In addition, other double mutants showed no effect on plant development, namely: *phe1::agl45*, *phe1::agl28*, *phe2::agl45*, *phe2::agl28*, *agl40::agl45*, *agl40::agl28* and *agl45::agl28*. These data suggested that this set of transcription factor insertion alleles needed to be in a higher order complex to show a phenotype.

6.3.4 All triple mutants had no seed phenotype except the *phe1::agl40::agl45* complex

The double mutants showed no effect on seed development including final seed weight. Therefore, the strategy was to create a higher order complex (triple) between the targeted genes. Four ‘complexes’ were created following the same protocol as described above, two with *phe2::agl40* and the others with *phe1::agl40* homozygous double mutants.

The triple homozygous mutants, namely: *phe2::agl40::agl45*, *phe2::agl40::agl28* and *phe1::agl40::agl28*, showed no phenotypic changes in the overall plant phenotype, such as stem length, leaf size, or flower size, or endosperm development and final seed weight. In addition, the *agl40/agl40::agl45/agl45::agl28/agl28* triple homozygous mutant showed a similar phenotype to the wild type (data not shown). In contrast, despite extensive efforts, plants with the genotype *phe1::agl40::agl45* were not recovered from self-pollinated *PHE1/phe1::AGL40/agl40::AGL45/agl45* plants. There were several potential explanations for the absence of the genotype, including pollen, ovule or seed lethality.

Considering the time available, only one closely related genotype, namely *phe1/phe1::agl40/agl40::AGL45/agl45*, was chosen for investigating the reason behind the failure in recovering triple homozygous plants. Also, because an abnormal seed phenotype

was detected in *phe1/phe1::agl40/agl40::AGL45/agl45* plants, we did not need to do any reciprocal crosses with the wild type to test pollen and ovule viability.

Whilst plants with the closely related genotype *phe1/phe1::agl40/agl40::AGL45/agl45* were vegetatively normal, in contrast the siliques produced by self pollination contained seeds of both a wild type appearance and abnormally small and brown seed-like structures (Figure 6.23b). Microscopy using a seed clearing protocol revealed that these structures contained both an embryo and endosperm (Figure 6. 23d-k), confirming them as seeds rather than unfertilized ovules. The abnormal seeds contained both a strikingly smaller endosperm (58% smaller in embryo sac area) and under-developed or misshapen embryos. The embryos in the abnormal seeds were abnormally developed, such as unbent cotyledons at a late embryogenesis stage compared to the normal embryo. Also, some of the abnormal seeds collapsed at the late heart or torpedo embryo stage. A possible explanation for this is that the endosperm is a nutrient sink that is supplied by the seed parent and nourishes the embryo during embryogenesis or germination, so successful embryogenesis therefore requires the development of the endosperm (Lopes and Larkins, 1993). Also, the endosperm has a role in supporting the early growth of the embryo until sufficient reserves are stored in the cotyledons (Marinos, 1970; Raghavan, 1986). Therefore, any failure in endosperm development will cause a failure or death of the embryo, so the potential cause of the misshapen embryo phenotype was the failure in endosperm development at an early stage of seed development.

We speculated above that there were several potential explanations for the absence of the genotype, including pollen, ovule or seed lethality. The presence of the aborted seeds in the silique rule out any possibility of a pollen lethality effect. Pollen lethality would not produce aborted seeds because the wild-type pollen would fertilize all of the ovules and produce normal seeds in the silique. Another suggestion would be ovule lethality effect; this would leave empty positions in the seed pod. Since the small white structures were confirmed as seeds because they contained an embryo and endosperm, this possibility is also eliminated. Therefore, the most likely explanation is that the 3:1 ratio of seed lethality was caused by the inheritance of the three insertion alleles. On examination of the mature dry siliques, it was found that they contained a mixture of seeds with a ratio of 75%:25%; normal to abortive seeds (Figure 6.23). The *MEDEA* gene is a maternally imprinted gene; its mutant (*mea*) produced 50% aborted seeds which collapsed and did not germinate in self-fertilised heterozygote plants (Grossniklaus *et al.*, 1998). This ratio (1:1; defective:

normal seeds) was consistent with a gametophytic control of the defect, because half of the haploid gametophytes received the mutant allele. In *mea/MEA* plants, heterozygous embryos were aborted if the mutant allele was derived from the female but they developed normally if they were derived from the male (Grossniklaus *et al.*, 1998). The *phe1/agl40/agl45* triple mutant could not have caused an effect on the seeds like the imprinted gene *MEDEA*, because the abortion frequency observed was 25% whereas the effect of the imprinted gene would be 50% seed abortion in the progeny; therefore, this possibility was eliminated.

The abortion frequency of about 25% was consistent with proposition that triple *phe1::agl40::agl45* mutants cause seed lethality. Although the embryos in the abnormal seeds found in the selfed pods of *phe1/phe1::agl40/agl40::AGL45/agl45* plants could not be genotyped, a reasonable assumption was that these seeds represented the missing *phe1/phe1::agl40/agl40::agl45/agl45* genotype. This finding suggests that the three transcription factors *PHE1*, *AGL40* and *AGL45* are required redundantly for successful seed development.

The *sep1-3* complex provides a good example of functional redundancy within the MADS box gene family, which is involved in the patterning of the floral meristem (Pelaz *et al.*, 2000). Single *sep* mutants showed wild-type phenotypes, consistent with the possibility of functional redundancy mutants, whilst the triple mutant (*sep1/sep2/sep3*) resulted in dramatic homeotic changes to flower development and resulted in flowers completely composed of sepals. Therefore, *SEP1/2/3* are a class of organ-identity genes required for the development of petals, stamens and carpels (Pelaz *et al.*, 2000).

Unlike single *sep* mutants, single *agl40* and *agl45* mutants showed no phenotypic changes, whereas the *phe1* single mutant showed a small seed phenotype, although the seeds were still viable (Chapter 4). However, similar to the *sep1-3* triple mutant, the triple *phe1/agl40/agl45* mutant showed dramatic homeotic changes to seed development (seed lethality).

Immink *et al.* (2009) suggested that combining protein interaction analyses and co-expression analyses provides complementary functional information about MADS box transcription factors, particularly when mutant phenotypes are missing due to redundancy or when the proteins are involved in multiple developmental processes. Figure 6.1, shows the protein-protein interaction between *PHE1*, *AGL40* and *AGL45*, either directly or

indirectly. This analysis suggests that these proteins have similar biological functions in plants. Therefore, we hypothesized that *PHE1*, *AGL40* and *AGL45* play the same role in seed development as growth promoters and that they represent functional redundancy mutants.

In summary, our findings on seed phenotypes and the final death of seeds suggest that the *phe1/phe1::agl40/agl40::agl45/agl45* triple homozygous mutant is seed lethal with a primary effect on endosperm development which leads to embryogenesis failure, causing seed death.

6.3.5 The *phe2*-centred quadruple mutant showed seed lethality

The *phe2* triple homozygous mutants discussed above showed no effect on seed development, especially final seed size. Therefore, the *phe2* quadruple mutant complex was created to test the effect on seed development of combining all four mutants, *phe2.1*, *agl40.1*, *agl45.1* and *agl28.1*, in a complex as described in Section 6.2.6. Despite extensive efforts, plants with the genotype *phe2::agl40::agl45::agl28* were not recovered from self-pollinated *PHE2/phe2::AGL40/agl40::AGL45/agl45::AGL28/agl28* plants. There were several potential explanations for the absence of the genotype, including pollen, ovule or seed lethality, as discussed earlier in Section 6.3.4.

Only one of the closely related genotypes (*phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28*) was selected for seed phenotype analyses because of the limited time available for this project. The *phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28* plants showed no vegetative phenotypic changes, such as stem length, leaf size or flower size compared to the *phe2* triple homozygous complexes and the wild type; however, there were differences in seed phenotype (Figure 6.25a-b). Although plants with the closely related genotype *phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28* were vegetatively normal, the siliques produced by self-pollination contained seeds of both a wild-type appearance and abnormally small and white seed-like structures (Figure 6.25a). Microscopy using a seed clearing protocol revealed that these structures contained both an embryo and endosperm (Figure 6.26), confirming them as seeds rather than unfertilized ovules. The white seed was smaller in size and contained an unstructured amorphous endosperm with no material or cells and an abnormal irregularly shaped embryo. In addition, there was no obvious chalazal endosperm in the seeds at all of the tested developmental stages (3-9 DAP), and at maturity the seeds had collapsed and become shrivelled (Figure 6.25b).

Further investigations using confocal microscopy confirmed the seed phenotypes observed by seed clearing (Figure 6.27). The abnormal seeds were relatively small at all stages and exhibited obvious abnormalities in both embryo and endosperm development from the earliest time point tested (3 DAP). In comparison to a normal seed phenotype, the abnormal seed embryo consisted of an irregularly shaped cluster of cells of a similar size (Figure 6.27q-t), indicating that embryo development arrested at an early stage. In addition, the embryo sac of the abnormal seed contained amorphous material and no normal endosperm nuclei or cells (Figure 6.27u-x).

The relative severity of the endosperm phenotype compared to that of the embryo suggests that the primary reason for seed abortion was a failure in endosperm development. In this case, the abnormal appearance of the embryo and its failure to progress would be secondary consequences of endosperm failure. Our results are consistent with reports showing that the size and developmental maturity of the endosperm in seed development appears to be a major determinant of seed survival and size (Boisnard-Lorig *et al.*, 2001; Garcia *et al.*, 2005). In addition, the endosperm is a nutrient sink that is supplied by the seed parent and nourishes the embryo during embryogenesis; successful embryogenesis therefore requires the development of the endosperm (Marinos, 1970; Raghavan, 1986; Lopes and Larkins, 1993). Any failure in endosperm development will therefore cause embryo lethality as observed here.

In Section 6.3.4 we discussed several potential explanations for the absence of the genotype, including pollen, ovule or seed lethality. Similar to the *phe1/agl40/AGL45* triple complex the *phe2/agl40/agl45/AGL28* quadruple complex produced aborted seeds in a silique, ruling out any possibility of pollen lethality. Similarly, ovule lethality would have left empty positions in the seed pod. Since the small white structures were confirmed as seeds because they contained an embryo and endosperm, this possibility was also eliminated.

Therefore, the most likely explanation is that the 3:1 ratio of seed lethality was caused by the inheritance of the three insertion alleles. Consistent with this proposition, the mature dry siliques contained a mixture of seeds with a ratio of 75%:25%; normal to abortive seeds (Figure 6.25c), suggesting that the aborted seeds likely represented the quadruple mutant (*phe2/phe2::agl40/agl40::agl45/agl45::agl28/agl28*) progeny that could not be recovered from selfed *phe2/phe2::agl40/agl40::agl45/agl45::AGL28/agl28* plants.

We discussed the case of the *phe1/agl40/agl45* triple mutants which may have a functional redundancy similar to the *sep1/sep2/sep3* mutants (Section 6.3.4). Further to this, the single mutants of *phe2*, *agl40*, *agl45* and *agl28* showed no phenotypic changes when compared to the wild type. However, they showed dramatic changes in seed development when as a quadruple complex (*phe2/agl40/agl45/agl28*). Therefore, it is likely that *PHE2*, *AGL40*, *AGL45* and *AGL28* are functionally redundant participants in the same biological process in plant development as the *PHE1/AGL40/AGL45* complex and *SEP1-3*.

Day *et al.* (2008) reported that *PHE2* is another endosperm-specific gene which has 72% homology to *PHE1* at the amino acid level. Analysis of the *PHE2* promoter-GUS gene in transgenic plants confirmed that the expression of *PHE2* is largely equivalent to *PHE1* expression during wild-type seed development. The author also suggested that *PHE1* and *PHE2* had a similar regulation pattern and perhaps a redundant function in seed development.

In summary, our findings suggest that the quadruple mutant *phe2/phe2::agl40/agl40::agl45/agl45::agl28/agl28* has a lethal effect on seed development, with the primary effect potentially being within the endosperm. However, it is unclear how *PHE2*, *AGL40*, *AGL45* and *AGL28* act to affect embryo and endosperm development.

In conclusion, based on expression studies, mutant analysis, and protein-protein interaction studies, we hypothesize that the type I MADS box genes play important roles in early seed development. Also, we can potentially alter seed size by manipulating the expression of these genes, thus providing a route to improvements in crop species.

6.3.6 Future work

We were successful in investigating the role in seed development of the candidate genes *PHE1*, *PHE2*, *AGL40*, *AGL45* and *AGL28* as triple and quadruple complexes. However, the exact mechanisms by which these genes participate in endosperm development remains unknown and will therefore require further investigation. One possible fruitful avenue would be to use RT-PCR to determine the expression profile of the individual candidate genes of the quadruple mutants in single mutants. It was recently found that in the absence of a functional copy of *PHE1*, *PHE2* expression increases by more than fivefold (Walia *et al.*, 2009). This will help us to understand how these genes work and to determine what role and level of control they have within the complex to regulate endosperm/seed development. Furthermore, the exact identity of the endosperm- and embryo-like structures

in the aborted seeds could be investigated further by assaying for embryo- or endosperm-specific gene expression, using either reporter genes or RT-PCR.

The mutations in individual genes were in various ecotypes for e.g. Col-0, *Ler* and Ws-4. Even though we had controls where the different ecotypes were combined, it is still possible that the interaction between individual genes might subtly vary between ecotypes. It is therefore recommended that all the different mutations be in one ecotype. This can be achieved by backcrossing the individual mutations to the desired ecotype for several generations.

Chapter 7: General discussion

Several studies have been conducted to change seed size in plants by altering endosperm growth and differentiation with the aim of increasing seed yield. Scott *et al.* (1998) carried out the first study using interploidy crosses in *Arabidopsis thaliana* (*A. thaliana*). This study revealed that different seed sizes could be generated by making reciprocal crosses between plants at different ploidy levels. The endosperm normally has a 2maternal:1paternal (2m:1p) genome ratio which results from a normal 2x X 2x cross; however, where a tetraploid (4x) plant is used as a pollen parent the resulting parental genome ratio in the endosperm is 2m:2p. This ‘paternal excess’ endosperm ratio leads to endosperm over-growth and increased seed size. The reciprocal ‘maternal excess’ cross (4x X 2x) results in an endosperm with a 4m:1p genome ratio, endosperm under-growth and a reduction in mature seed mass. As discussed in Chapter 1, these parent-of-origin effects point both to the participation of uniparentally silenced, imprinted genes in the process of endosperm growth and development and that these genes encode positive or negative growth regulators. .

The primary aim of the work presented in this thesis was to identify genes with the potential to engineer endosperm growth and development and therefore to increase seed size and yield in crop plants. Since genes that are either subject directly or indirectly to imprinting were likely to act to increase or decrease endosperm proliferation, the transcriptome of seeds generated by interploidy crosses was potentially a rich source of useful genes. The work reported in this thesis aimed to identify such genes, and in particular genes capable of causing increased endosperm proliferation.

7.1 The candidate genes are positively associated with endosperm proliferation

The Scott laboratory had previously used a microarray approach to analyse the expression of seed genes in different interploidy crosses relative to the balanced 2x X 2x cross (Tiwari *et al.*, 2010). Genes displaying large changes in relative expression, either increasing or decreasing, were listed as candidate genes potentially responsible for the observed differences in endosperm size and seed size generated in the various crosses (see Chapter 1). This list contained a large number of candidate genes that could have been verified by further experiment. However, given the limited resources, a smaller but coherent subset of genes was sought for verification.

The candidate genes namely: *PHERES 1 (PHE1)*, *PHERES 2 (PHE2)*, *AGL40*, *AGL45*, *AGL62* and *AGL28* are highly expressed in paternalized endosperm (See Table 3.2). These genes belong to the same family of MADS box transcription factor proteins members of which are known to be key regulators of several plant development processes (Alvarez-Buylla *et al.*, 2000a; Alvarez-Buylla *et al.*, 2000b; Parenicova *et al.*, 2003). Other interesting genes, belonging to different families were also selected namely: *CYCD4;1*, *CYCD4;2*, *GAI*, *GA20-oxidase 5*, *GA-regulated*, *AT5G46950*, *CKX2*, *MAPK10*, and *E2L2*. Again, all were highly expressed in paternalized endosperm (See Table 3.2). One main objective of the current project was to assess the role of these genes in endosperm development principally using two complementary molecular genetics approaches: KI (Chapter 3) and KO (Chapters 4, 5 and 6).

7.2 The potential of candidate genes to alter seed size

7.2.1 The outcomes of the KI strategy

In general, the constitutive over-expression of the candidate genes from the CaMV 35S promoter did not provide informative data. Transgenic *A. thaliana* plants carrying the transgenes: *35S::PHE2*, *35S::AGL40* and *35S::AT5G46950* were indistinguishable from wild type plants. In contrast, *35S::E2L2*, *35S::MAPK10* and *35S::CKX2* constructs produced plants with obvious abnormal phenotypes (Section 3.2.2). However, these phenotypes were difficult to interpret as resulting from the over-expression of genes that normally function as cellular growth promoters, as was predicted from their elevated expression in paternal excess crosses. Others have noted that constitutive expression of foreign genes may cause abnormalities in the host plants, such as sterility, retarded development, abnormal morphology, yield penalties, altered grain composition or transgene silencing (Cai *et al.*, 2007). For this reason, strong tissue-specific or inducible promoters that limit gene expression to the required tissue or at a particular time may solve this type of problem. Consequently, the present study switch to creating transgenic over-expression lines in which the candidate genes were expressed from an endosperm specific promoter.

Endosperm-specific over-expression of the *PHE2* transcription factor led to an increase in seed weight of approximately 12%. Fertility as measured by seed number/silique in plants containing the *pPER::PHE2* transgene was normal suggesting that expression of the transgene was directly responsible for the observed increase in seed size (Section 3.2.3.1).

Similarly, *pPER::GAI* plants had a larger seed size with an increase in seed fertility compared to wild type. This again suggests that high level of *GAI* expression in endosperm had the potential to increase both seed size and the number of seed per pod (Section 3.2.3.3). The mechanism by which endosperm-specific over expression of *GAI* could increase seed number per pod is not understood.

Whilst over-expression of *PHE2* and *GAI* genes showed potential for increasing seed size and yield, modifying the expression of other genes did not have the desired results. *A. thaliana* plants harbouring *pPER::AGL40*, *pPER::AGL45*, *pPER::AGL62*, *pPER::GA20OX5*, *pPER::GA-regulated*, *pPER::CYCD4;1* or *pPER::CYCD4;2* constructs had no discernable impact on seed (Chapters 3 and 5)

7.2.2 The outcomes of the KO strategy

7.2.2.1 Loss of *PHE1* expression reduces endosperm growth

PHE1 is an imprinted gene that is expressed paternally and suppressed maternally (Kohler *et al.*, 2005); therefore according to the parental conflict model (Haig and Westoby, 1989; 1991) *PHE1* should encode growth promoter. Although, *phe1.1* previously used in studies (Kohler *et al.*, 2005; Josefsson *et al.*, 2006), the work in Chapter 4 provided the first study of *phe1.1* phenotype and its role in endosperm development.

phe1.1 plants produced ovules containing a relatively small embryo sac, a reduced number of peripheral endosperm nuclei, and relatively a small seed size (Section 4.2.2). This phenotype was associated with a significant reduction in total seed yield. The data strongly suggest that *PHE1* encodes a growth promoter that is required for the correct development of the endosperm. This is consistent with the KI over-expression data discussed above.

In addition, the results obtained from restricted and unrestricted 4x *phe1.1* interploidy crosses showed that an extra copy of paternal *PHE1* potentially contributes to seed death in interploidy crosses. There was a significant reduction in seed abortion (5%) in the 2x X 4x *phe1.1* compared to 2x X 4x control cross (13%). This suggests that lethal endosperm over-proliferation that causes seed abortion in interploidy crosses in *A. thaliana* (Scott *et al.*, 1998; Adams *et al.*, 2000; Comai *et al.*, 2000; Bushell *et al.*, 2003; Dilkes and Comai, 2004; Dilkes *et al.*, 2008) requires paternal *PHE1* expression. In contrast, the 4x *phe1.1* X 2x cross showed a dramatic increase in the frequency of shrivelled seed (16%) over the 4x X 2x control cross (6%). This effect was unexpected since *PHE1* is preferentially

expressed from the paternal genome (Kohler *et al.*, 2005) consistent with its role as an imprinted growth promoter in our study on *phe1* loss of function mutant. Clearly, however, maternal expression of *PHE1* plays a role in seed development. There have been reports of the paternal genome remaining transcriptionally inactive during the early phase of seed development following fertilisation (Vielle-Calzada *et al.*, 2000). The effect of the maternal *PHE1* could thus be either direct, by providing PHE1 protein during the earliest stages of endosperm development when the expression from the paternal *PHE1* allele may be limited (Vielle-Calzada *et al.*, 2000), or indirect which involves maintaining a balance of expression with its interacting or controlling partners. If some amount of PHE1 protein is indeed required during the early stages of development, then lack of expression during this period may reduce endosperm size below a critical level in a proportion of seeds resulting in increased seed abortion. These findings suggest that *PHE1* has an important role in controlling seed abortion in both 2x X 4x and 4x X 2x crosses in *A. thaliana*.

Our study of 4x *phe1.1* interploidy crosses yields results that are in line with another study on the role of *PHE1* in affecting seed survival in intraspecific crosses in a dosage dependent manner (Josefsson *et al.*, 2006). The authors suggested that over-expression of *PHE1* from the maternal genome contributed to hybrid seed death in *A. thaliana* X *A. arenosa* crosses and that knocking out maternal *PHE1* rescued hybrid seed survival. Although the studies are not directly comparable; 2x *A. thaliana* X 2x *A. arenosa* is lethal whereas the comparable reciprocal 4x X 2x cross is not, *PHE1* is biallelic in the *A. thaliana* X *A. arenosa* cross but remains maternally imprinted in the *Arabidopsis* interploidy crosses (Josefsson *et al.* 2006). Nevertheless, both studies give support to the differential dosage hypothesis, which suggests that imprinting, developed to control the relative dosage of regulatory factors in the endosperm (Dilkes and Comai, 2004).

In conclusion, our study on the phenotypic characterisation of the *phe1.1* loss of function mutant suggests that *PHE1* plays a critical role in endosperm growth and development and functions as a growth promoter.

7.2.2.2 AGL62 controls endosperm cellularisation

As described in Chapter 5, approximately 25% of the seeds in an *AGL62/agl62* heterozygous silique suffered early seed abortion, strongly suggesting that homozygous *agl62/agl62* allele is lethal (recessive seed lethality). The *AGL62* loss-of-function developing seeds displayed very obvious developmental differences to wild type seed from

2 DAP onward. The most striking features of these seeds were embryo arrest at the 8-cell stage and very early cellularisation of the endosperm (Figures 5.3 and 5.4). These findings are in agreement with data published by Kang *et al.* (2008). The authors found that *agl62* mutant seeds showed three types of defects: precocious endosperm cellularisation, reduced number of endosperm nuclei and abnormal embryo development. *AGL62* is reportedly expressed exclusively in the endosperm (Table 3.2). Its expression is high in the syncytial endosperm and declines at the onset of cellularisation; it is also expressed from both parental alleles (Sections 5.2.1.2 and 5.2.1.3), indicating that it is not imprinted. *AGL62* has been implicated as an important controller of endosperm cellularisation (Kang *et al.*, 2008). Among the evidence, is that in the *fis* mutants where the endosperm fails to cellularise, the level of *AGL62* transcript remains high until the seed collapses and aborts (Kang *et al.*, 2008). Our findings along with those of Kang *et al.* (2008), suggest that *AGL62* is a growth promoter that functions to prolong the proliferation phase of endosperm development at the expense of differentiation (cellularisation).

7.2.2.3 MADS box transcription factor genes showed functional redundancy in controlling endosperm development

Amongst the genes that were highly expressed in the paternal excess crosses (large seeds) were a set of MADS box transcription factor genes namely: *PHE1*, *PHE2*, *AGL62*, *AGL40*, *AGL45* and *AGL28*. *AGL62* was the only gene in this group with a strong KO phenotype and was therefore considered separately in Chapter 5. KO lines for the other genes did not have detectable phenotypes except for *PHE1* that showed a small seed size (Chapter 4). In a yeast two hybrid experiment elucidating interaction between the MADS box transcription factor family (de Folter *et al.*, 2005), the proteins of these genes have been shown to interact with each other. This has also been demonstrated in an *in silico* generated interaction map in Chapter 6. Loss of function phenotype of individual genes of the group were combined as double mutants but again showed no obvious differences to the wild type (Sections 6.3.2 and 6.3.3). In all the various mutant combinations that were studied (Chapter 6), seed lethality phenotype was restricted to the triple (*phe1/agl40/agl45*) and quadruple (*phe2/agl40/agl45/agl28*) homozygous complexes. *phe1/agl40/agl45* seeds had under-developed or misshapen embryos, 58% smaller embryo sac area and collapsed seeds (Figure 6.23). Roughly 25% of the seeds had the abortive and abnormal shape phenotype. These failed to germinate hence we could not recover any plants with *phe1/agl40/agl45* mutant allele combination. Because the genes are expressed in the endosperm, the effect of the complex was predicted to be mainly in the endosperm with abnormalities in the embryo

being an indirect effect. The homozygous white seeds of the *phe2/agl40/agl45/agl28* quadruple mutant combination were smaller in size and showed irregular shaped embryo and unstructured amorphous endosperm (Figures 6.26 and 6.27). Again, roughly 25% seeds were shrivelled at maturity, and because we could not recover any plant with the *phe2/agl40/agl45/agl28* genotype we concluded that the abortion/abnormality was due to simultaneous loss of function of all the four genes in the complex.

Our results are consistent with reports showing that the size and developmental maturity of the endosperm in seed development appears to be a major determinant of seed survival and size (Boisnard-Lorig *et al.*, 2001; Garcia *et al.*, 2005). The endosperm is a nutrient sink that is supplied by the seed parent and nourishes the embryo during embryogenesis; successful embryogenesis requires a proper development of the endosperm (Marinos, 1970; Raghavan, 1986; Lopes and Larkins, 1993). Any failure in endosperm development will therefore cause embryo lethality as observed in the mutant complexes. We also suggest that these genes have an essential role in endosperm development as a growth promoter, they act in a complex and that a degree of functional redundancy exists. A good example of functional redundancy within MADS box family is the *sep1-3* complex, which is involved in the patterning of the floral meristem (Pelaz *et al.*, 2000; Chapter 6). Our results on seed phenotypes and seed abortion suggest that the triple mutant *phe1/phe1::agl40/agl40::agl45/agl45* and the quadruple mutant *phe2/phe2::agl40/agl40::agl45/agl45::agl28/agl28* have a lethal effect on seed development, with the primary effect potentially being within the endosperm. However, it is unclear how PHE1, PHE2, AGL40, AGL45 and AGL28 act to affect embryo and endosperm development.

In conclusion, expression studies (Tiwari *et al.*, 2010; Chapter 3), mutant analysis (Chapter 6), and protein-protein interaction studies (de Folter *et al.*, 2005), demonstrate that type I MADS box genes play important roles in early seed development. This knowledge is potentially useful in manipulating seed size and therefore seed yield, for example by pointing to the need to over-express appropriate combinations of MADS box genes within the endosperm. The necessity of a multi-gene approach is supported by the results of the over-expression studies described in Chapter 3, in which over-expression of single genes in the endosperm had little or no effect on endosperm or seed size.

The study outlined in the thesis has thus opened new areas of research which when taken forward will provide a better understanding of genes and the networks controlling

endosperm development, crosstalk between the different seed components and provide new insights and tools to manipulate seed size and yield.

References

- Adams, S., Vinkenoog, R., Spielman, M., Dickinson, H.G., Scott, R.J., 2000. Parent-of-origin effects on seed development in *Arabidopsis thaliana* require DNA methylation. *Development* **127**, 2493-2502.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., Ecker, J.R., 2003. Genoe-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653-657.
- Alvarez-Buylla, E.R., Liljegren, S.J., Pelaz, S., Gold, S.E., Burgeff, C., Ditta, G.S., Vergara-Silva, F., Yanofsky, M.F., 2000a. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant Journal* **24**, 457-466.
- Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., de Pouplana, L.R., Martinez-Castilla, L., Yanofsky, M.F., 2000b. Ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5328-5333.
- Andreasson, E., Ellis, B., 2010. Convergence and specificity in the *Arabidopsis* MAPK nexus. *Trends in Plant Science* **15**, 106-113.
- Arumuganathan, K., Earle, E., 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* **9**, 208-218.
- Baroux, C., Spillane, C., Grossniklaus, U., 2002. Evolutionary origins of the endosperm in flowering plants. *Genome Biology* **3** (9), 1026.1-1026.5.
- Baroux, U., Gagliardini, V., Page, D.R., Grossniklaus, U., 2006. Dynamic regulatory interactions of Polycomb group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis*. *Genes and Development* **20**, 1081-1086.
- Bartel, D.P., 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297.
- Becker, A., Theissen, G., 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution* **29**, 464-489.
- Benfey, P.N., Ren, L., Chua, N.H., 1989. The CAMV 35S Enhancer contains at least 2 domains which can confer different developmental and tissue-specific expression patterns. *EMBO Journal* **8**, 2195-2202.
- Berger, F., 2003. Endosperm: the crossroad of seed development. *Current Opinion in Plant Biology* **6**, 42-50.
- Berger, F., Grini, P.E., Schnittger, A., 2006. Endosperm: an integrator of seed growth and development. *Current Opinion in Plant Biology* **9**, 664-670.
- Berger, F., Hamamura, Y., Ingouff, M., Higashiyama, T., 2008. Double fertilization - caught in the act. *Trends in Plant Science* **13**, 437-443.

- Boisnard-Lorig, C., Colon-Carmona, A., Bauch, M., Hodge, S., Doerner, P., Bancharel, E., Dumas, C., Haseloff, J., Berger, F., 2001.** Dynamic analyses of the expression of the HISTONE::YFP fusion protein in *Arabidopsis* show that syncytial endosperm is divided in mitotic domains. *Plant Cell* **13**, 495-509.
- Bouche, N., Bouchez, D., 2001.** *Arabidopsis* gene knockout: phenotypes wanted. *Current Opinion in Plant Biology* **4**, 111-117.
- Braselton, J.P., Wilkinson, M.J., Clulow, S.A., 1996.** Feulgen staining of intact plant tissues for confocal microscopy. *Biotechnic and Histochemistry* **71**, 84-87.
- Brown, R.C., Lemmon, B.E., Nguyen, H., 2003.** Events during the first four rounds of mitosis establish three developmental domains in the syncytial endosperm of *Arabidopsis thaliana*. *Protoplasma* **222**, 167-174.
- Brown, R.C., Lemmon, B.E., Nguyen, H., Olsen, O.A., 1999.** Development of endosperm in *Arabidopsis thaliana*. *Sexual Plant Reproduction* **12**, 32-42.
- Bushell, C., Spielman, M., Scott, R.J., 2003.** The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *Plant Cell* **15**, 1430-1442.
- Cai, M., Wei, J., Li, X.H., Xu, C.G., Wang, S.P., 2007.** A rice promoter containing both novel positive and negative *cis*-elements for regulation of green tissue-specific gene expression in transgenic plants. *Plant Biotechnology Journal* **5**, 664-674.
- Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., Peacock, W.J., 1997.** Fertilization-independent seed development in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4223-4228.
- Cheng, W.H., Taliercio, E.W., Chourey, P.S., 1996.** The *Miniature1* seed locus of Maize encodes a cell wall invertase required for normal development of endosperm and maternal cells in the pedicel. *Plant Cell* **8**, 971-983.
- Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M.M., Colombo, L., 2008.** *AGL23*, a type I MADS-box gene that controls female gametophyte and embryo development in *Arabidopsis*. *Plant Journal* **54**, 1037-1048.
- Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R., Reynolds, S.H., Stevens, Y., Byers, B., 2000.** Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* **12**, 1551-1568.
- Cornejo, M.J., Luth, D., Blankenship, K.M., Anderson, O.D., Blechl, A.E., 1993.** Activity of a Maize ubiquitin promoter in transgenic rice. *Plant Molecular Biology* **23**, 567-581.
- Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., Pirrotta, V., 2002.** *Drosophila* Enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185-196.
- Day, R.C., Herridge, R.P., Ambrose, B.A., Macknight, R.C., 2008.** Transcriptome analysis of proliferating *Arabidopsis* endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. *Plant Physiology* **148**, 1964-1984.
- de Folter, S., Immink, R.G.H., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B., Angenent, G.C., 2005.** Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. *Plant Cell* **17**, 1424-1433.
- Dilkes, B.P., Comai, L., 2004.** A differential dosage hypothesis for parental effects in seed development. *Plant Cell* **16**, 3174-3180.

- Dilkes, B.P., Spielman, M., Weizbauer, R., Watson, B., Burkart-Waco, D., Scott, R.J., Comai, L.,** 2008. The maternally expressed *WRKY* transcription factor *TTG2* controls lethality in interploidy crosses of *Arabidopsis*. *PLoS Biology* **6**, 2707-2720.
- Duvick, D.N.,** 1999. Heterosis: Feeding people and protecting natural resources. American Society of Agronomy, Madison, USA.
- Duvick, D.N.,** 2005. Genetic progress in yield of United States Maize (*Zea mays L.*). *Maydica* **50**, 193-202.
- Edwards, K., Johnstone, C., Thompson, C.,** 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research* **19**, 1349-1349.
- Food and Agriculture Organisation for United Nations (FAO),** 2002. Reducing poverty and hunger: the critical role of financing for food, agriculture and rural development. <ftp://ftp.fao.org/docrep/fao/oo3/y6265e/y6265.pdf>.
- Food and Agriculture Organisation for United Nations (FAO),** 2004. Special report: FAO/WFP crop and food supply assessment mission to Sudan. <http://www.fao.org/docrep/oo6/j1505e/j1505e00.pdf>.
- Food and Agriculture Organisation for United Nations (FAO),** 2007. FAOSTAT, faostat.fao.org/site/567/default.aspx#ancor.
- Food and Agriculture Organisation for United Nations (FAO),** 2009a. How to feed the world in 2050. <http://www.fao.org/wsfs/forum2050/wsfs-forum/en/>.
- Food and Agriculture Organisation for United Nations (FAO),** 2009b. The State of Food Insecurity in the World. <http://www.fao.org/publications/sofi/en/>.
- Feilner, T., Hultschig, C., Lee, J., Meyer, S., Immink, R.G.H., Koenig, A., Possling, A., Seitz, H., Beveridge, A., Scheel, D., Cahill, D.J., Lehrach, H., Kreutzberger, J., Kersten, B.,** 2005. High throughput identification of potential *Arabidopsis* mitogen-activated protein kinases substrates. *Molecular and Cellular Proteomics* **4**, 1558-1568.
- Finger, R.,** 2010. Evidence of slowing yield growth - The example of Swiss cereal yields. *Food Policy* **35**, 175-182.
- Floyd, S.K., Friedman, W.E.,** 2000. Evolution of endosperm developmental patterns among basal flowering plants. *International Journal of Plant Sciences* **161**, S57-S81.
- Focks, N., Benning, C.,** 1998. *wrinkled1*: A novel, low-seed-oil mutant of *Arabidopsis* with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiology* **118**, 91-101.
- Friedman, W.E.,** 1998. The evolution of double fertilization and endosperm: an "historical" perspective. *Sexual Plant Reproduction* **11**, 6-16.
- Friedman, W.E., Madrid, E.N., Williams, J.H.,** 2008. Origin of the fittest and survival of the fittest: Relating female gametophyte development to endosperm genetics. *International Journal of Plant Sciences* **169**, 79-92.
- Friedman, W.E., Ryerson, K.C.,** 2009. Reconstructing the ancestral female gametophyte of angiosperms: Insights from Amborella and other ancient lineages of flowering plants. *American Journal of Botany* **96**, 129-143.

- Friedman, W.E., Williams, J.H.**, 2004. Developmental evolution of the sexual process in ancient flowering plant lineages. *Plant Cell* **16**, S119-S132.
- Gambin, B.L., Borras, L.**, 2010. Resource distribution and the trade-off between seed number and seed weight: a comparison across crop species. *Annals of Applied Biology* **156**, 91-102.
- Garcia, D., Fitz Gerald, J.N., Berger, F.**, 2005. Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *Plant Cell* **17**, 52-60.
- Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jurgens, G., Berger, F.**, 2003. *Arabidopsis haiku* mutants reveal new controls of seed size by endosperm. *Plant Physiology* **131**, 1661-1670.
- Gehring, M., Huh, J.H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., Fischer, R.L.**, 2006. DEMETER DNA glycosylase establishes *MEDEA* Polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**, 495-506.
- Grini, P.E., Jurgens, G., Hulskamp, M.**, 2002. Embryo and endosperm development is disrupted in the female gametophytic capulet mutants of *Arabidopsis*. *Genetics* **162**, 1911-1925.
- Grossniklaus, U., Vielle-Calzada, J.-P., Hoepfner, M.A., Gagliano, W.B.**, 1998. Maternal Control of Embryogenesis by *MEDEA*, a Polycomb Group Gene in *Arabidopsis*. *Science* **280**, 446-450.
- Guilley, H., Dudley, R.K., Jonard, G., Balázs, E., Richards, K.E.**, 1982. Transcription of cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. *Cell* **30**, 763-773.
- Guittou, A.E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.E., Grossniklaus, U., Berger, F.**, 2004. Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in *Arabidopsis thaliana*. *Development* **131**, 2971-2981.
- Haig, D., Westoby, M.**, 1989. Parent-specific gene-expression and the triploid endosperm. *American Naturalist* **134**, 147-155.
- Haig, D., Westoby, M.**, 1991. Genomic imprinting in endosperm- Its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **333**, 1-13.
- Harbour, J.W., Dean, D.C.**, 2000. Chromatin remodelling and Rb activity. *Current Opinion in Cell Biology* **12**, 685-689.
- Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.-T., Benfey, P.N.**, 2000. The *SHORT-ROOT* gene controls radial patterning of the *Arabidopsis* root through radial signalling. *Cell* **101**, 555-567.
- Henderson, I.R., Jacobsen, S.E.**, 2007. Epigenetic inheritance in plants. *Nature* **447**, 418-424.
- Huh, J.H., Bauer, M.J., Hsieh, T.-F., Fischer, R.**, 2007. Endosperm gene imprinting and seed development. *Current Opinion in Genetics and Development* **17**, 480-485.
- Huh, J.H., Bauer, M.J., Hsieh, T.F., Fischer, R.L.**, 2008. Cellular programming of plant gene imprinting. *Cell* **132**, 735-744.
- Hughes, R.**, 2009. Determinants of seed size and yield in *Arabidopsis thaliana*. PhD thesis. Department of Biology and Biochemistry. University of Bath, Bath, UK.

- Immink, R.G.H., Tonaco, I.A.N., de Folter, S., Shchennikova, A., van Dijk, A.D.J., Busscher-Lange, J., Borst, J.W., Angenent, G.C., 2009.** *SEPALLATA3*: the 'glue' for MADS box transcription factor complex formation. *Genome Biology* **10** (2).
- Iwakawa, H., Shinmyo, A., Sekine, M., 2006.** *Arabidopsis CDKA1*, a *cdc2* homologue, controls proliferation of generative cells in male gametogenesis. *Plant Journal* **45**, 819-831.
- Jefferson, R.A., Kavanagh, T.A., Bevan, M.W., 1987.** GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO journal* **6**, 3901-3907.
- Jenuwein, T., Allis, C.D., 2001.** Translating the histone code. *Science* **293**, 1074-1080.
- Jofuku, K.D., Omidyar, P.K., Gee, Z., Okamuro, J.K., 2005.** Control of seed mass and seed yield by the floral homeotic gene *APETALA2*. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 3117-3122.
- Josefsson, C., Dilkes, B., Comai, L., 2006.** Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology* **16**, 1322-1328.
- Kang, I.H., Steffen, J.G., Portereiko, M.F., Lloyd, A., Drews, G.N., 2008.** The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* **20**, 635-647.
- Kermicle, J.L., Alleman, M., 1990.** Gametic imprinting in Maize in relation to the angiosperm life-cycle. *Development* **108**, 9-14.
- Kiniry, J.R., Wood, C.A., Spanel, D.A., Bockholt, A.J., 1990.** Seed weight response to decreased seed number in Maize. *Agronomy Journal* **82**, 98-102.
- Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., Fischer, R.L., 1999.** Imprinting of the *MEDEA* Polycomb gene in the *Arabidopsis* endosperm. *Plant Cell* **11**, 1945-1952.
- Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J.J., Goldberg, R.B., Fischer, R.L., 1999.** Control of fertilization-independent endosperm development by the *MEDEA* Polycomb gene *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4186-4191.
- Kohler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., Gruissem, W., 2003a.** *Arabidopsis* MS1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO Journal* **22**, 4804-4814.
- Kohler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., Grossniklaus, U., 2003b.** The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. *Genes and Development* **17**, 1540-1553.
- Kohler, C., Page, D.R., Gagliardini, V., Grossniklaus, U., 2005.** The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of *PHERES1* by parental imprinting. *Nature Genetics* **37**, 28-30.
- Kohler, C., Scheid, O.M., Erilova, A., 2010.** The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics* **26**, 142-148.
- Kono, A., Ohno, R., Umeda-Hara, C., Uchimiya, H., Umeda, M., 2006.** A distinct type of cyclin D, CYCD4;2, involved in the activation of cell division in *Arabidopsis*. *Plant Cell Reports* **25**, 540-545.
- Kono, A., Umeda-Hara, C., Adachi, S., Nagata, N., Konomi, M., Nakagawa, T., Uchimiya, H., Umeda, M., 2007.** The *Arabidopsis* D-type cyclin CYCD4 controls cell division in the stomatal lineage of the hypocotyl epidermis. *Plant Cell*, tpc.106.046763.

- Kosugi, S., Ohashi, Y., 2002.** E2Ls, E2F-like repressors of *Arabidopsis* that bind to E2F sites in a monomeric form. *Journal of Biological Chemistry* **277**, 16553-16558.
- Kranz, E., Kumlehn, J., 1999.** Angiosperm fertilisation, embryo and endosperm development in vitro. *Plant Science* **142**, 183-197.
- Lachner, M., Jenuwein, T., 2002.** The many faces of histone lysine methylation. *Current Opinion in Cell Biology* **14**, 286-298.
- Lachner, M., O'Sullivan, R.J., Jenuwein, T., 2003.** An epigenetic road map for histone lysine methylation. *Journal of Cell Science* **116**, 2117-2124.
- Lara-Nunez, A., de Jesus, N., Vazquez-Ramos, J.M., 2008.** Maize D4;1 and D5 cyclin proteins in germinating maize. Associated kinase activity and regulation by phytohormones. *Physiologia Plantarum* **132**, 79-88.
- Liljegren, S.J., Ditta, G.S., Eshed, H.Y., Savidge, B., Bowman, J.L., Yanofsky, M.F., 2000.** *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**, 766-770.
- Lin, B.Y., 1984.** Ploidy barrier to endosperm development in Maize. *Genetics* **107**, 103-115.
- Linkies, A., Graeber, K., Knight, C., Leubner-Metzger, G., 2010.** The evolution of seeds. *New Phytologist* **186**, 817-831.
- Lopes, M.A., Larkins, B.A., 1993.** Endosperm origin, development, and function. *Plant Cell* **5**, 1383-1399.
- Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., Peacock, W.J., Chaudhury, A.M., 1999.** Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 296-301.
- Luo, M., Dennis, E.S., Berger, F., Peacock, W.J., Chaudhury, A., 2005.** *MINISEED3 (MINI3)*, a *WRKY* family gene, and *HAIKU2 (IKU2)*, a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17531-17536.
- Luo, M., Luo, M.-Z., Buzas, D., Finnegan, J., Helliwell, C., Dennis, E.S., Peacock, W.J., Chaudhury, A., 2008.** UBIQUITIN-SPECIFIC PROTEASE 26 is required for seed development and the repression of *PHERES1* in *Arabidopsis*. *Genetics* **180**, 229-236.
- Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., Grossniklaus, U., Kohler, C., 2006.** Different Polycomb group complexes regulate common target genes in *Arabidopsis*. *EMBO Reports* **7**, 947-952.
- Makarevich, G., Villar, C.B.R., Erilova, A., Kohler, C., 2008.** Mechanism of *PHERES1* imprinting in *Arabidopsis*. *Journal of Cell Science* **121**, 906-912.
- Mariconti, L., Pellegrini, B., Cantoni, R., Stevens, R., Bergounioux, C., Cella, R., Albani, D., 2002.** The E2F family of transcription factors from *Arabidopsis thaliana*. *Journal of Biological Chemistry* **277**, 9911-9919.
- Marinos, N.G., 1970.** Embryogenesis of the pea (*Pisum sativum*). *Protoplasma* **71**, 227-233.
- Mascarenhas, J.P., 1989.** The male gametophyte of flowering plants. *Plant Cell* **1**, 657-664.
- Masubelele, N.H., Dewitte, W., Menges, M., Maughan, S., Collins, C., Huntley, R., Nieuwland, J., Scofield, S., Murray, J.A.H., 2005.** D-type cyclins activate division in the root apex to promote seed germination in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15694-15699.

- Menges, M., de Jager, S.M., Gruissem, W., Murray, J.A.H., 2005.** Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *Plant Journal* **41**, 546-566.
- Micheli, F., 2001.** Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends in Plant Science* **6**, 414-419.
- Miller, M.E., Chourey, P.S., 1992.** The Maize invertase-deficient *MINIATURE-1* seed mutation is associated with aberrant pedicel and endosperm development. *Plant Cell* **4**, 297-305.
- Moore, R.C., Grant, S.R., Purugganan, M.D., 2005.** Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. *Molecular Biology and Evolution* **22**, 91-103.
- Morgan, D.O., 1997.** Cyclin-dependent kinases: Engines, clocks, and microprocessors. *Annual Review of Cell and Developmental Biology* **13**, 261-291.
- Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., Simon, J.A., 2002.** Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197-208.
- Nam, J., Kim, J., Lee, S., An, G., Ma, H., Nei, M., 2004.** Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1910-1915.
- Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M., Lepiniec, L., 2002.** The *TRANSPARENT TESTA16* locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* **14**, 2463-2479.
- Ng, M., Yanofsky, M.F., 2001.** Function and evolution of the plant MADS-box gene family. *Nature Reviews Genetics* **2**, 186-195.
- Nguyen, H., Brown, R.C., Lemmon, B.E., 2000.** The specialized chalazal endosperm in *Arabidopsis thaliana* and *Lepidium virginicum* (Brassicaceae). *Protoplasma* **212**, 99-110.
- Nieuwland, J., Maughan, S., Dewitte, W., Scofield, S., Sanz, L., Murray, J.A.H., 2009.** The D-type cyclin CYCD4;1 modulates lateral root density in *Arabidopsis* by affecting the basal meristem region. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 22528-22533.
- Nowack, M.K., Grini, P.E., Jakoby, M.J., Lafos, M., Koncz, C., Schnittger, A., 2006.** A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. *Nature Genetics* **38**, 63-67.
- Odell, J.T., Nagy, F., Chua, N.H., 1985.** Identification of DNA-Sequences required for activity of the cauliflower Mosaic Virus-35S promoter. *Nature* **313**, 810-812.
- Ogawa, M., Hanada, A., Yamauchi, Y., Kuwalhara, A., Kamiya, Y., Yamaguchi, S., 2003.** Gibberellin biosynthesis and response during *Arabidopsis* seed germination. *Plant Cell* **15**, 1591-1604.
- Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B., Fischer, R.L., 1999.** Mutations in *FIE*, a WD Polycomb group gene, allow endosperm development without fertilization. *Plant Cell* **11**, 407-415.
- Ohto, M., Floyd, S.K., Fischer, R.L., Goldberg, R.B., Harada, J.J., 2009.** Effects of *APETALA2* on embryo, endosperm, and seed coat development determine seed size in *Arabidopsis*. *Sexual Plant Reproduction* **22**, 277-289.

- Ohto, M., Fischer, R.L., Goldberg, R.B., Nakamura, K., Harada, J.J., 2005.** Control of seed mass by *APETALA2*. Proceedings of the National Academy of Sciences of the United States of America **102**, 3123-3128.
- Olsen, O.A., 2004.** Nuclear endosperm development in cereals and *Arabidopsis thaliana*. Plant Cell **16**, S214-S227.
- Parenicova, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C., Colombo, L., 2003.** Molecular and phylogenetic analyses of the complete MADS-Box transcription factor family in *Arabidopsis*: New openings to the MADS world. Plant Cell **15**, 1538-1551.
- Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., Yanofsky, M.F., 2000.** B and C floral organ identity functions require *SEPALLATA* MADS-box genes. Nature **405**, 200-203.
- Pelloux, J., Rustérucci, C., Mellerowicz, E.J., 2007.** New insights into pectin methylesterase structure and function. Trends in Plant Science **12**, 267-277.
- Portereiko, M.F., Lloyd, A., Steffen, J.G., Punwani, J.A., Otsuga, D., Drews, G.N., 2006.** *AGL80* is required for central cell and endosperm development in *Arabidopsis*. Plant Cell **18**, 1862-1872.
- Raghavan, V., 1986.** Embryogenesis in Angiosperms: a development and experimental study. Cambridge University Press, Cambridge, UK.
- Ramirez-Parra, E., Lopez-Matas, M.A., Frundt, C., Gutierrez, C., 2004.** Role of an Atypical E2F transcription factor in the control of *Arabidopsis* cell growth and differentiation. Plant Cell **16**, 2350-2363.
- Reik, W., Dean, W., Walter, J., 2001.** Epigenetic reprogramming in mammalian development. Science **293**, 1089-1093.
- Ronemus, M.J., Galbiati, M., Ticknor, C., Chen, J.C., Dellaporta, S.L., 1996.** Demethylation-induced developmental pleiotropy in *Arabidopsis*. Science **273**, 654-657.
- Rossi, V., Locatelli, S., Lanzanova, C., Boniotti, M.B., Varotto, S., Pipal, A., Goralik-Schramel, M., Lusser, A., Gatz, C., Gutierrez, C., Motto, M., 2003.** A Maize histone deacetylase and retinoblastoma-related protein physically interact and cooperate in repressing gene transcription. Plant Molecular Biology **51**, 401-413.
- Sabelli, P.A., Larkins, B.A., 2009.** The development of endosperm in grasses. Plant Physiology **149**, 14-26.
- Sadras, V.O., 2007.** Evolutionary aspects of the trade-off between seed size and number in crops. Field Crops Research **100**, 125-138.
- Saurin, A.J., Shao, Z., Erdjument-Bromage, H., Tempst, P., Kingston, R.E., 2001.** A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. Nature **412**, 655-660.
- Schruff, M.C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., Scott, R.J., 2006.** The *AUXIN RESPONSE FACTOR 2* gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. Development **133**, 251-261.
- Schwartz, Y.B., Pirrotta, V., 2007.** Polycomb silencing mechanisms and the management of genomic programmes. Nature Review Genetics **8**, 9-22.
- Scott, R.J., Spielman, M., Bailey, J., Dickinson, H.G., 1998.** Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development **125**, 3329-3341.

- Shah, M., Strong, M., F.,** 2002. Food in the 21st century: from science to sustainable agriculture. World Bank. <http://books.google.co.uk>.
- Shore, P., Sharrocks, A.D.,** 1995. The MADS box family of transcription factors. *European Journal of Biochemistry* **229**, 1-13.
- Somerville, C., Koornneef, M.,** 2002. Timeline - A fortunate choice: the history of *Arabidopsis* as a model plant. *Nature Reviews Genetics* **3**, 883-889.
- Sorensen, M.B., Chaudhury, A.M., Robert, H., Bancharel, E., Berger, F.,** 2001. Polycomb group genes control pattern formation in plant seed. *Current Biology* **11**, 277-281.
- Spielman, M., Vinkenoog, R., Dickinson, H.G., Scott, R.J.,** 2001. The epigenetic basis of gender in flowering plants and mammals. *Trends in Genetics* **17**, 705-711.
- Steffen, J.G., Kang, I.H., Portereiko, M.F., Lloyd, A., Drews, G.N.,** 2008. AGL61 interacts with AGL80 and is required for central cell development in *Arabidopsis*. *Plant Physiology* **148**, 259-268.
- Steinbrecher, R.A.,** 2002. The CaMV 35S Promoter, Government and Corporate Scientific Incompetence: Failure to assess the safety of GM crops. ECONEXUS. <http://www.econexus.info/publications.html>.
- Sundaresan, V.,** 2005. Control of seed size in plants. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17887-17888.
- Sunilkumar, G., Mohr, L., Lopata-Finch, E., Emani, C., Rathore, K.S.,** 2002. Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. *Plant Molecular Biology* **50**, 463-474.
- Terada, R., Shimamoto, K.,** 1990. Expression of CAMV35S-GUS gene in transgenic rice plants *Molecular and General Genetics* **220**, 389-392.
- Tiwari, S., Spielman, M., Day, R.C., Scott, R.J.,** 2006. Proliferative phase endosperm promoters from *Arabidopsis thaliana*. *Plant Biotechnology Journal* **4**, 393-407.
- Tiwari, S., Spielman, M., Schulz, R., Oakey, R.J., Kelsey, G., Salazar, A., Zhang, K., Pennell, R., Scott, R.J.,** 2010. Transcriptional profiles underlying parent-of-origin effects in seeds of *Arabidopsis thaliana*. *BMC Plant Biology* **10** (72), 1-22.
- Ungru, A., Nowack, M.K., Reymond, M., Shirzadi, R., Kumar, M., Biewers, S., Grini, P.E., Schnittger, A.,** 2008. Natural variation in the degree of autonomous endosperm formation reveals independence and constraints of embryo growth during seed development in *Arabidopsis thaliana*. *Genetics* **179**, 829-841.
- Vaillant, I., Paszkowski, J.,** 2007. Role of histone and DNA methylation in gene regulation. *Current Opinion in Plant Biology* **10**, 528-533.
- Venable, D.L.,** 1992. Size-number trade-offs and the variation of seed size with plant resource status. *The American Naturalist* **140**, 287.
- Vielle-Calzada, J.-P., Baskar, R., Grossniklaus, U.,** 2000. Delayed activation of the paternal genome during seed development. *Nature* **404**, 91-94.

- Vielle-Calzada, J.-P., Thomas, J., Spillane, C., Coluccio, A., Hoepfner, M.A., Grossniklaus, U., 1999.** Maintenance of genomic imprinting at the *Arabidopsis medea* locus requires zygotically DDM1 activity. *Genes and Development* **13**, 2971-2982.
- Villar, C.B.R., Erilova, A., Makarevich, G., Trosch, R., Kohler, C., 2009.** Control of *PHERES1* imprinting in *Arabidopsis* by direct tandem repeats. *Molecular Plant* **2**, 654-660.
- Vinkenoog, R., Scott, R.J., 2001.** Autonomous endosperm development in flowering plants: how to overcome the imprinting problem?. *Sexual Plant Reproduction* **14**, 189-194.
- Vinkenoog, R., Spielman, M., Adams, S., Fischer, R.L., Dickinson, H.G., Scott, R.J., 2000.** Hypomethylation promotes autonomous endosperm development and rescues postfertilization lethality in *fi* mutants. *Plant Cell* **12**, 2271-2282.
- Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J., Comai, L., 2009.** Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Current Biology* **19**, 1128-1132.
- Wellmer, F., Riechmann, J.L., 2005.** Gene network analysis in plant development by genomic technologies. *International Journal of Developmental Biology* **49**, 745-759.
- Werner, T., Kollmer, I., Bartrina, I., Holst, K., Schmulling, T., 2006.** New insights into the biology of cytokinin degradation. *Plant Biology* **8**, 371-381
- Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H., Schmulling, T., 2003.** Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell* **15**, 2532-2550.
- West, A.G., Shore, P., Sharrocks, A.D., 1997.** DNA binding by MADS-box transcription factors: A molecular mechanism for differential DNA bending. *Molecular and Cellular Biology* **17**, 2876-2887.
- World Food Programme (WFP), 2009.** World hunger. <http://www.wfp.org/hunger>.
- Williamson, J.D., Hirschwyncott, M.E., Larkins, B.A., Gelvin, S.B., 1989.** Differential Accumulation of a Transcript Driven by the CaMV 35S Promoter in Transgenic Tobacco. *Plant Physiology* **90**, 1570-1576.
- Xiao, W., Brown, R.C., Lemmon, B.E., Harada, J.J., Goldberg, R.B., Fischer, R.L., 2006.** Regulation of seed size by hypomethylation of maternal and paternal genomes. *Plant Physiology* **142**, 1160-1168.
- Xiao, W., Gehring, M., Choi, Y., Margossian, L., Pu, H., Harada, J.J., Goldberg, R.B., Pennell, R.I., Fischer, R.L., 2003.** Imprinting of the *MEA* Polycomb gene is controlled by antagonism between MET1 Methyltransferase and DME Glycosylase. *Developmental Cell* **5**, 891-901.
- Yang, N.-S., Christou, P., 1990.** Cell type specific expression of a CaMV 35S-GUS gene in transgenic soybean plants. *Developmental Genetics* **11**, 289-293.
- Yoo, S.K., Lee, J.S., Ahn, J.H., 2006.** Overexpression of *AGAMOUS-LIKE 28* (*AGL28*) promotes flowering by upregulating expression of floral promoters within the autonomous pathway. *Biochemical and Biophysical Research Communications* **348**, 929-936.

Zhang, J.Z., 2003. Overexpression analysis of plant transcription factors. *Current Opinion in Plant Biology* **6**, 430-440.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., Jacobsen, S.E., 2007. Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biology* **5**, e129.